Production Lead Time in Exponential Serial Lines: Analysis and Control

Semyon M. Meerkov and Chao-Bo Yan

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109-2122

Abstract

Production lead time (LT) is the average time a part spends in the system, being processed and waiting for processing. In the previous work, we have developed methods for analysis and control of LT in production lines with machines obeying the Bernoulli reliability model. While this model is applicable to some assembly operations, it is not applicable to operations, where the downtime is much longer than the machine cycle time, e.g., turning, boring, milling, drilling, grinding, etc. Therefore, the current paper is devoted to analysis and control of LT in serial lines with machine reliability models having wide applicability, i.e., exponential, Weibull, gamma, and log-normal. More specifically, we develop methods for analysis as well as open- and closed-loop control of LT in synchronous serial lines with exponential machines and then extend some of these results to asynchronous lines and non-Markovian reliability models.

Keywords: Production systems, Lead time, Exponential and non-exponential reliability models, Open- and closed-loop raw material release control.

1 Introduction

Production lines are typically managed to maximize their throughput. In some cases, this leads to excessively long production lead time (LT), i.e., the average time a part spends in the system, being processed and waiting for processing. Long lead time may be unacceptable for economic and quality reasons. These considerations call for a different management paradigm: *operate production systems so that the desired lead time is ensured, while the throughput is maximized.*

This constrained optimization problem, which we referred to as the *lead time control* (LTC) problem, has been addressed in [1] and [2] for serial and cellular lines, respectively. Operationally, the approach was based on "throttling" the raw material release rate so that the desired performance is attained. Mathematically, both papers provided a solution assuming that the machines obey the Bernoulli reliability model [3], whereby each machine is up during a cycle time with probability p and down with probability $1 - p$. While this model is appropriate for assembly operations, where the machine downtime is typically short, it is not applicable to machining operations with downtime much longer than the machine cycle time. Along with its limited applicability, the Bernoulli reliability model does not provide a possibility for investigating effects of the machines' up- and downtime on the LT behavior.

The current paper, while also using the throttling approach, is intended to eliminate the shortcomings of [1] and [2] by considering machine reliability models, where the downtime may be much longer than machine cycle time, i.e., exponential, Weibull, gamma, and log-normal. Specifically, we address in details the analysis and open- and closed-loop control of LT in synchronous exponential serial lines (i.e., the lines with machines having identical cycle time and exponentially distributed up- and downtimes) and then provide upper bounds on LT for asynchronous exponential lines and synchronous lines with Weibull, gamma, and log-normal reliability models. Since the bounds for all non-exponential cases considered here are the same, we conjecture that they hold for all continuous models of machine reliability.

As far as the literature review is concerned, publications on production lead time can be classified into three groups. The first one considers the lead time as a function of the dispatch (rather than release) rule [4–10]. Dispatch rules indicate which job must be selected for processing at a given workcenter. The main result here is that, under a wide range of conditions, jobs with the shortest processing time should be selected first in order to minimize the lead time. The second group addresses the issue of feedback control of raw material release. The main control strategies considered are kanban [11–23] and CONWIP [24–35]. However, this literature does not provide methods for selecting parameters of these control strategies (i.e., the number of kanbans or the limit of CONWIP), which would lead to the desired lead time. The third group consists of papers [1] and [2] mentioned above, which provide formulas for the lead time as a function of Bernoulli machine parameters and use these formulas for solving the LTC problem in Bernoulli lines. As indicated above, the current paper is intended to advance this research by considering systems with continuous machine reliability models of practical importance.

The outline of this paper is as follows: Section 2 introduces the model and the problems addressed. Sections 3-5 consider synchronous exponential lines. Section 6 is devoted to extensions to asynchronous exponential and non-exponential lines. The conclusions and topics for future research are given in Section 7. All proofs are included in the Appendix.

2 Modeling and Problems Addressed

Consider a serial line shown in Figure 2.1, where the circles represent the machines and the open rectangles are the buffers. While $m_1, m_2, \ldots, m_M$ and $b_1, b_2, \ldots, b_{M-1}$ are the usual producing machines and work-in-process buffers, respectively, m_0 represents the raw material release machine and b_0 raw material buffer (to indicate this, m_0 and b_0 are shown in gray). Controlling the efficiency of the release machine, m_0, one can control the availability of raw material in the system and, thus, the lead time.

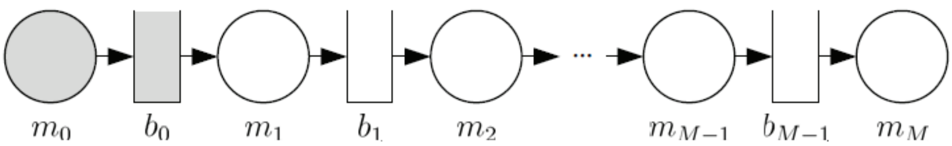


Figure 2.1: Serial production line with a release machine

To formalize this model, we introduce the following assumptions:

(i) The system consists of M producing machines, $m_1, m_2, \ldots, m_M$, a release machine, m_0, $M-1$ work-in-process buffers, $b_1, b_2, \ldots, b_{M-1}$, and a raw material buffer, b_0.

(ii) Each machine is characterized by its cycle time, τ_i (in min), $i = 0, 1, \ldots, M$. If cycle times of all machines (including the release machine) are identical, the system is called *synchronous*; otherwise, it is *asynchronous*. While in the asynchronous case, τ_i, $i = 1, 2, \ldots, M$, are fixed, τ_0 is free and can be selected at will.

(iii) In addition, each machine is characterized by its *reliability model*, i.e., continuous random variables that define its up- and downtime. If these distributions are exponential, i.e., defined by the breakdown rate λ_i and repair rate μ_i, $i = 0, 1, \ldots, M$, (both in 1/min), the line is called *exponential*; otherwise, it is *non-exponential*. While for the producing machines, λ_i and μ_i, $i = 1, 2, \ldots, M$, are fixed, for the release machine, λ_0 and μ_0 are design parameters that can be selected at will.

(iv) Each buffer is of infinite capacity.

(v) The flow model [3] is assumed, (i.e., infinitesimal quantity of parts, produced during an infinitesimal time interval, are transferred to and from the buffers). A machine is starved, if the buffer in front of it is empty; m_0 is never starved. Machine failures are time-dependent [3], i.e., a machine can be down even if it is starved.

Assumption (iv) is introduced, on one hand, to simplify the presentation, and, on the other hand, to reflect the fact that the LTC problem is of particular importance for systems with practically unlimited storage, e.g., with no hardware-constrained buffers, so that many parts can be stored between each pair of consecutive operations. Assumption (v) is introduced for technical reasons: it permits a precise formulation of the equations describing the systems at hand.

Let $T_{up,i}$ and $T_{down,i}$ denote the average up- and downtime of the machines, $i = 0, 1, \ldots, M$. Then the machine *efficiency* for any continuous reliability model is (see [3]):

$$e_i := \frac{T_{up,i}}{T_{up,i} + T_{down,i}}, \; i = 0, 1, \ldots, M, \tag{2.1}$$

and its *throughput in isolation* (i.e., when the machine is not starved) is

$$TP_{isol,i} := \frac{T_{up,i}}{\tau_i(T_{up,i} + T_{down,i})}, \ i = 0, 1, \ldots, M. \tag{2.2}$$

Since for exponential machines, $T_{up,i} = \frac{1}{\lambda_i}$ and $T_{down,i} = \frac{1}{\mu_i}$,

$$e_i = \frac{\mu_i}{\lambda_i + \mu_i} \text{ and } TP_{isol,i} = \frac{\mu_i}{\tau_i(\lambda_i + \mu_i)}, \ i = 0, 1, \ldots, M. \tag{2.3}$$

Clearly, to obtain meaningful results, it should be assumed that $e_0 < e_i$, $i = 1, 2, \ldots, M$, for synchronous lines or $TP_{isol,0} < TP_{isol,i}$, $i = 1, 2, \ldots, M$, for asynchronous ones (otherwise, *LT* becomes unbounded).

In the case of finite buffer capacity, a method for evaluating the throughput (*TP*) and work-in-process in each buffer (WIP_i, $i = 0, 1, \ldots, M$) of serial lines defined above is given in [3]. In the current paper, we modify this method for the case of infinite buffers and address the following problems:

1. Develop an analytical method for evaluating *LT* in synchronous exponential lines as a function of the producing machines and the release machine parameters.

2. For synchronous exponential lines with given (λ_i, μ_i), $i = 1, 2, \ldots, M$, develop a method for solving the open-loop LTC problem, i.e., for selecting (λ_0, μ_0) so that *LT* takes the desired value, while maximizing *TP*.

3. For synchronous exponential lines, develop a method for solving the closed-loop LTC problem, which would allow to maintain the desired *LT* even if the parameters of the producing machines, (λ_i, μ_i), $i = 1, 2, \ldots, M$, are not known precisely.

4. Extend the solutions of the above open- and closed-loop LTC problems to the case when the raw material release is deterministic (e.g., once-per-hour or once-per-shift), rather than random (once-per-cycle).

5. Generalize the above results to asynchronous exponential and non-exponential lines.

Solutions of problems 1-4 are given in Sections 3-5 and Section 6 provides a solution of problem 5.

3 Analysis of Lead Time in Synchronous Exponential Lines

Below and in the subsequent section, we first address the case of identical producing machines, where the results are especially transparent and instructive, and then generalize them to the non-identical machine case.

3.1 Identical producing machines

3.1.1 General properties

Proposition 3.1 *Consider a synchronous exponential serial line defined by assumptions (i)-(v). Assume that all producing machines are identical, i.e.,* $\lambda_i = \lambda$, $\mu_i = \mu$, $i = 1, 2, \ldots, M$, *and the release machine is less efficient than the producing machines, i.e.,* $e_0 < e$. *Then, an estimate of the lead time (in min) is given by*

$$\widehat{LT} = M\tau + \left[\frac{e_0}{\mu_0} + (2M-1)\frac{e}{\mu}\right]\left(\frac{1-e}{e-e_0}\right). \tag{3.1}$$

Proof: See the Appendix.

The accuracy of this estimate was evaluated by simulating exponential lines with identical machines and with parameters M, e, e_0, T_{down}, and $T_{down,0}$ selected randomly and equiprobably from the following sets:

$$M \in [3, 10],\ e \in [0.7, 0.99],\ e_0 \in [0.7e, 0.99e],\ T_{down} \in [10\text{min}, 100\text{min}],\ T_{down,0} = T_{down}. \tag{3.2}$$

For each line, thus constructed, the analysis was carried out for two τ's: $\tau = 0.5$ min and $\tau = 5$ min. The total of 1000 lines have been simulated using the following procedure: For each line, in addition to a warm-up period of $2,000,000$ minutes, the simulation was carried out for $22,000,000$ minutes; 20 repetitions of this procedure were carried out to evaluate LT. This simulation procedure results in a 95% confidence interval of ±0.87% of LT for both $\tau = 0.5$ min and $\tau = 5$ min. The accuracy of (3.1) was quantified by $\epsilon_{LT} = \frac{|\widehat{LT}-LT|}{LT} \times 100\%$. As a result, we obtained: For $\tau = 0.5$ min, the smallest and the largest errors were 0.0025% and 8.97%, respectively, and the average error was 2.17%; for $\tau = 5$ min, the smallest and the largest errors were 0.0007% and 7.21%, respectively, and the average error was 1.99%. Based on these data and recognizing that machine parameters on the factory floor are rarely known with accuracy better than ±5%, we conclude that estimate (3.1) is precise enough for the lead time analysis and control.

Expression (3.1) leads to the following conclusions:

- For fixed e, shorter up- and downtimes of the producing machines (i.e., larger λ and μ) lead to smaller $\widehat{LT}$.
- Similarly, for fixed e_0, shorter up- and downtimes of the release machine (i.e., larger λ_0 and μ_0) lead to smaller $\widehat{LT}$.
- $\widehat{LT}$ is monotonically increasing in M, hyperbolically increasing as $e_0 \to e$, and is an affine function of τ with the slope M.
- As $e \to 1$, $\widehat{LT}$ tends to its minimum value, $M\tau$.

To further characterize the behavior of $\widehat{LT}$, introduce the following parametrization:

$$\rho := \frac{e_0}{e}, \tag{3.3}$$

$$\widehat{lt} := \frac{\widehat{LT}}{M\tau}. \tag{3.4}$$

We refer to $0 < \rho < 1$ as the *relative workload* imposed on the system and to $\widehat{lt} > 1$ as the *relative (dimensionless) lead time*, i.e., the lead time in units of the smallest possible lead time. In terms of these parameters, (3.1) becomes

$$\widehat{lt} = 1 + \frac{1}{\tau}\Big(\frac{\rho}{M\mu_0} + \frac{2M-1}{M\mu}\Big)\Big(\frac{1-e}{1-\rho}\Big). \tag{3.5}$$

Clearly, in addition to M, e, and τ, the relative lead time, $\widehat{lt}$, depends on the release machine efficiency, e_0 (through ρ) and on its downtime (through μ_0). However, in the limit as M tends to infinity, the dependency on μ_0 disappears:

$$\widehat{lt}_\infty := \lim_{M\to\infty} \widehat{lt} = 1 + \frac{2}{\mu\tau}\Big(\frac{1-e}{1-\rho}\Big). \tag{3.6}$$

This is convenient for the LTC problem, since for long lines, only e_0 would have to be selected, rather than μ_0 as well. To evaluate how well $\widehat{lt}_\infty$ approximates $\widehat{lt}$, consider

$$\Delta = \frac{\widehat{lt}_\infty - \widehat{lt}}{\widehat{lt}_\infty}. \tag{3.7}$$

Then, using (3.5) and (3.6), it is possible to show that

$$\Delta = \frac{\frac{1}{M\tau}\Big(\frac{1}{\mu} - \frac{\rho}{\mu_0}\Big)\Big(\frac{1-e}{1-\rho}\Big)}{1 + \frac{2}{\mu\tau}\Big(\frac{1-e}{1-\rho}\Big)},$$

and, if $\mu_0 \geqslant \mu$ (i.e., $T_{down,0} \leqslant T_{down}$),

$$0 < \Delta < \frac{1}{2M}. \tag{3.8}$$

Thus, $\widehat{lt}_\infty \geqslant \widehat{lt}$ and the difference between them decreases hyperbolically in M; therefore, (3.6) can be used as a relatively tight bound of (3.5) in serial lines with, say, ten or more machines (leading to errors less than 5%).

From (3.6) follows another observation: If $\mu\tau = 2e$, then (3.6) becomes

$$\widehat{lt}_\infty = \frac{e^{-1} - \rho}{1-\rho}. \tag{3.9}$$

This expression is exactly the same as the expression for $\widehat{lt}$ in Bernoulli serial lines (see [1]), with the Bernoulli machine efficiency p substituted by the exponential machine efficiency e. Hence, from (3.6) and (3.9), we conclude:

- If $\mu\tau = 2e$, then $\widehat{lt}_\infty$ in exponential lines equals $\widehat{lt}$ in Bernoulli lines with the same producing machine efficiency, i.e., $p = e$.

- If $\mu\tau < 2e$, then $\widehat{lt}_\infty$ in exponential lines is larger than $\widehat{lt}$ in Bernoulli lines with $p = e$. Since $\mu\tau < 2e$ implies that, for any $0 < e < 1$, $T_{down} > \frac{\tau}{2}$ (which is practically always the case), we conclude that $\widehat{lt}$ *in exponential lines* (quantified by $\widehat{lt}_\infty$) *is generically larger than* $\widehat{lt}$ *in Bernoulli lines*.

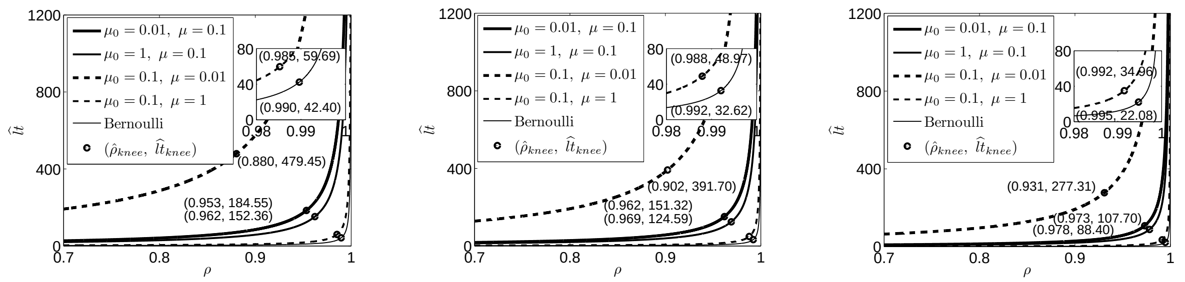
3.1.2 Knee-type behavior

Figure 3.1 illustrates the behavior of $\widehat{lt}$ given by (3.5) as a function of ρ for $M = 10$, $\tau = 1$min and several values of e, μ, and μ_0; the Bernoulli case, i.e., when $\mu\tau = 2e$, is also shown for comparison purposes. All curves in this figure have a "knee" beyond which $\widehat{lt}$ grows extremely fast. It is of interest to characterize "safe" release rates, i.e., the release rates below the knee. To accomplish this, consider the $(\rho, \widehat{lt})$-plane, where a unit interval of ρ-axis corresponds to $A > 1$ units of $\widehat{lt}$-axis (in Figure 3.1, $A = 4000$). Introduce the *scaling ratio*, α, defined by

$$\alpha := \frac{1}{A} \tag{3.10}$$

and recall that the *curvature*, κ, of a twice differentiable function, $f(x)$, is given by (see [36])

$$\kappa(f(x)) = \frac{\left|f''_{xx}\right|}{(1 + f'^2_x)^{\frac{3}{2}}}. \tag{3.11}$$



(a) $e = 0.7$ (b) $e = 0.8$ (c) $e = 0.9$

Figure 3.1: Relative lead time, $\widehat{lt}$, as a function of relative workload, ρ, and machine parameters (for $M = 10$, $\tau = 1$ min)

Definition 3.1 *The knee,* $\hat{\rho}_{knee}$, *of* $\widehat{lt}$ *on the* $(\rho, \widehat{lt})$*-plane with the scaling ratio* α *is the point on* $[0, 1)$ *at which the curvature of* $\alpha\widehat{lt}(\rho)$ *reaches its maximum.*

Proposition 3.2 *Under the assumptions of Proposition 3.1,*

$$\hat{\rho}_{knee} = 1 - \sqrt{\frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu}\Big)(1-e)} \tag{3.12}$$

and

$$\lim_{M\to\infty} \hat{\rho}_{knee} = 1 - \sqrt{\frac{2\alpha}{\mu\tau}(1-e)}. \tag{3.13}$$

Proof: See the Appendix.

The pairs $(\hat{\rho}_{knee}, \widehat{lt}(\hat{\rho}_{knee}))$ are indicated in Figure 3.1 by black dots. Thus, releasing raw material with the rate

$$e_0 < e\Big(1 - \sqrt{\frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu}\Big)(1-e)}\Big), \tag{3.14}$$

or, as $M \to \infty$,

$$e_0 < e\Big(1 - \sqrt{\frac{2\alpha}{\mu\tau}(1-e)}\Big), \tag{3.15}$$

results in $\widehat{lt}$ below the knee. Observe that, as it follows from (3.12) and (3.13), the position of the knee shifts to the right (i.e., larger release rates become safe) if the producing machine efficiency is increased or the up- and downtime of all machines are decreased.

Note that in practice, the position of the knee is referred to as the "sweet point". Thus, (3.12) and (3.13) provide an analytical tool for selecting raw material release rates that ensure system operation at the sweet point.

3.2 Non-identical producing machines

3.2.1 General properties

Proposition 3.3 *Consider a synchronous exponential serial line defined by assumptions (i)-(v). Assume that the release machine is less efficient than the producing machines, i.e.,* $e_0 < \min_{1\leqslant i\leqslant M} e_i$*. Then, an estimate of the lead time (in min) is given by*

$$\widehat{LT} = M\tau + \sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big). \tag{3.16}$$

Proof: See the Appendix.

Clearly, this expression reduces to (3.1) if all producing machines are identical. Also, it is obvious that the qualitative properties of (3.1) hold for (3.16) as well. For instance, for fixed e_i, $i = 0, 1, \dots, M$, shorter up- and downtimes lead to shorter $\widehat{LT}$, and $\widehat{LT}$ tends to its minimum (i.e., $M\tau$) as $e_i \to 1$, $i = 1, 2, \dots, M$.

The accuracy of this estimate has been evaluated by simulating serial lines with non-identical machines and with M, e_i, $T_{down,i}$, $i = 1, 2, \dots, M$, e_0, and $T_{down,0}$ selected randomly and equiprobably from the sets

$$\begin{aligned} &M \in [3, 10],\ e_i \in [0.8, 0.99],\ i = 1, 2, \dots, M,\ e_0 \in [0.8 \min_{1 \leqslant i \leqslant M} e_i,\ 0.99 \min_{1 \leqslant i \leqslant M} e_i], \\ &T_{down,0} \in [10\text{min}, 100\text{min}],\ T_{down,i} \in [T_{down,0}, 1.1T_{down,0}],\ i = 1, 2, \dots, M. \end{aligned} \tag{3.17}$$

For each line, the analysis was carried out with $\tau = 0.5$ min and $\tau = 5$ min. The total of 1000 lines have been investigated using the simulation procedure outlined in Subsection 3.1, again using $\epsilon_{LT} = \frac{|\widehat{LT} - LT|}{LT} \times 100\%$ as the measure of accuracy. The results turned out to be less precise than in the identical producing machines case. Namely, for $\tau = 0.5$ min, the smallest and largest errors were 0.0138% and 19.65%, respectively, and the average error was 3.94%; for $\tau = 5$ min, the smallest and largest errors were 0.0037% and 19.00%, respectively, and the average error was 3.71%. However, when e_i's, $i = 0, 1, \dots, M$, were selected from sets $e_i \in [0.9, 0.99]$, $i = 1, 2, \dots, M$, and $e_0 \in [0.9 \min_{1 \leqslant i \leqslant M} e_i,\ 0.99 \min_{1 \leqslant i \leqslant M} e_i]$, the accuracy was similar to that of the identical producing machines case: for $\tau = 0.5$ min, the smallest and largest errors were 0.0002% and 9.71%, respectively, and the average error was 1.42%; for $\tau = 5$ min, the smallest and largest errors were 0.0002% and 9.54%, respectively, and the average error was 1.32%.

To further investigate $\widehat{LT}$ defined by (3.16), introduce a modified relative load factor

$$\rho_{max} := \frac{e_0}{e_{min}}, \tag{3.18}$$

where $e_{min} = \min_{1 \leqslant i \leqslant M} e_i$, while keeping the relative lead time, $\widehat{lt}$, as in (3.4). Although reducing (3.16) to an expression for $\widehat{lt}(\rho_{max})$ leads to a complicated formula, the following upper bounds present a clearer picture:

Proposition 3.4 *Under the assumptions of Proposition 3.3,*

$$\widehat{lt} \leqslant \overline{\overline{lt}} := 1 + \frac{1}{\tau}\left(\frac{\rho_{max}}{M\mu_0} + \frac{2M-1}{M\mu_{min}}\frac{e_{max}}{e_{min}}\right)\left(\frac{1-e_{min}}{1-\rho_{max}}\right), \tag{3.19}$$

where $e_{min} = \min\limits_{1\leqslant i\leqslant M} e_i$, $e_{max} = \max\limits_{1\leqslant i\leqslant M} e_i$, *and* $\mu_{min} = \min\limits_{1\leqslant i\leqslant M} \mu_i$. *Also, in the limit as* $M \to \infty$,

$$\overline{\overline{lt}}_\infty := \lim_{M\to\infty} \overline{\overline{lt}} = 1 + \left(\frac{2}{\tau\mu_{min}}\frac{e_{max}}{e_{min}}\right)\left(\frac{1-e_{min}}{1-\rho_{max}}\right). \tag{3.20}$$

Proof: See the Appendix.

Note that if the producing machines are identical, these expressions reduce to (3.5) and (3.6), respectively. Also, all qualitative properties of (3.5) and (3.6) hold for (3.19) and (3.20) as well. For instance, (3.20) does not depend on μ_0, while (3.19) does. Finally, the rate of convergence of (3.19) to (3.20) for $\mu_0 \geqslant \mu_{min}$, as quantified by (3.7), is also $\frac{1}{2M}$, so that the following chain of inequalities take place:

$$\widehat{lt}(\rho_{max}) \leqslant \overline{\overline{lt}}(\rho_{max}) \leqslant \overline{\overline{lt}}_\infty(\rho_{max}). \tag{3.21}$$

This implies that if the release rate e_0 is selected so that the bound (3.20) satisfies the desired lead time, LT_d, the system performance will be at least as good as LT_d.

3.2.2 Knee-type behavior

Similar to the identical machine case, function $\widehat{lt}(\rho_{max})$ exhibits a knee-type behavior. This is illustrated in Figure 3.2 (solid curves) for the following three 10-machine lines:

$$\begin{aligned}
L_1:\ \boldsymbol{e} &= [0.75, 0.63, 0.73, 0.68, 0.75, 0.70, 0.73, 0.69, 0.67, 0.66],\\
\boldsymbol{T_{down}} &= [13.87, 23.64, 16.06, 20.83, 13.02, 23.96, 17.57, 27.20, 27.07, 21.87],\\
L_2:\ \boldsymbol{e} &= [0.73, 0.78, 0.85, 0.72, 0.74, 0.75, 0.75, 0.82, 0.76, 0.75],\\
\boldsymbol{T_{down}} &= [10.31, 24.94, 18.90, 28.64, 19.32, 18.37, 26.92, 20.50, 14.05, 23.44],\\
L_3:\ \boldsymbol{e} &= [0.90, 0.97, 0.96, 0.93, 0.96, 0.93, 0.87, 0.86, 0.87, 0.91],\\
\boldsymbol{T_{down}} &= [24.54, 16.19, 26.77, 21.36, 17.41, 24.06, 20.93, 18.90, 23.89, 22.43],
\end{aligned} \tag{3.22}$$

where $\boldsymbol{e}$ and $\boldsymbol{T_{down}}$ are the vectors of producing machine efficiency and downtime, respectively. The parameters of the producing machines of these lines have been selected randomly and equiprobably from the following sets: $T_{down,i} \in [10\text{min}, 30\text{min}]$ and $e_i \in [0.9e, 1.1e]$, $i = 1, 2, \ldots, M$, with $e = 0.7$ for line L_1, 0.8 for line L_2, and 0.9 for line L_3. For all three lines, the cycle time τ was selected as 1 min and the release machine downtime as 10 min.

While it seems impossible to quantify the position of the knee of $\widehat{lt}(\rho_{max})$, it is possible to lower-bound it by considering the knee of $\bar{\bar{\widehat{lt}}}(\rho_{max})$ or $\bar{\bar{\widehat{lt}}}_\infty(\rho_{max})$. The behavior of these functions is also shown in Figure 3.2 (by dashed and dash-dot curves, which practically overlay each other). The position of their knees can be quantified as follows:

Proposition 3.5 *Under the assumptions of Proposition 3.3, the knees of* $\bar{\bar{\widehat{lt}}}$ *and* $\bar{\bar{\widehat{lt}}}_\infty$ *are given, respectively, by*

$$\bar{\bar{\rho}}_{knee}(\bar{\bar{\widehat{lt}}}) = 1 - \sqrt{\frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu_{min}}\frac{e_{max}}{e_{min}}\Big)(1 - e_{min})} \tag{3.23}$$

and

$$\bar{\bar{\rho}}_{\infty,knee}(\bar{\bar{\widehat{lt}}}_\infty) = \lim_{M\to\infty} \bar{\bar{\rho}}_{knee}(\bar{\bar{\widehat{lt}}}) = 1 - \sqrt{\frac{2\alpha e_{max}}{\tau\mu_{min}e_{min}}(1 - e_{min})}. \tag{3.24}$$

Proof: See the Appendix.

A lower bound on the knee of $\widehat{lt}(\rho_{max})$ is given by the following:

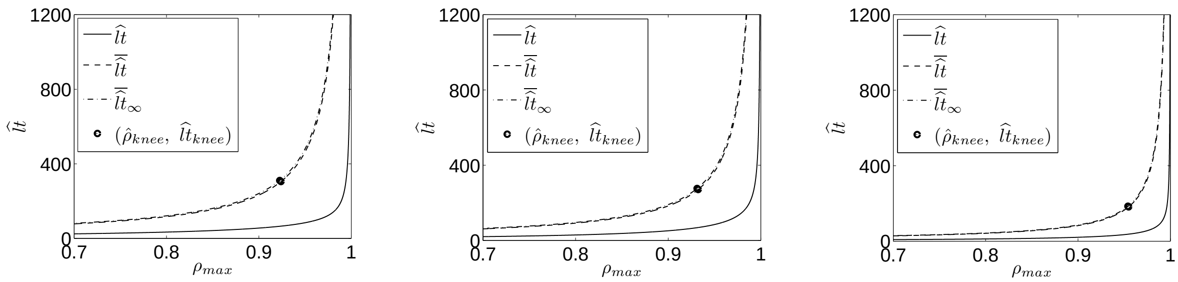
Proposition 3.6 *Under the assumptions of Proposition 3.3 and with* $\mu_0 \geqslant \mu_{min}$,

$$\hat{\rho}_{knee}(\widehat{lt}) \geqslant \bar{\bar{\rho}}_{\infty,knee}(\bar{\bar{\widehat{lt}}}_\infty). \tag{3.25}$$

Proof: See the Appendix.

The knees of $\bar{\bar{\widehat{lt}}}_\infty(\rho_{max})$ are shown in Figure 3.2 by black dots. Thus, releasing raw material with the load factor $\rho_{max} \leqslant \bar{\bar{\rho}}_{\infty,knee}$ ensures a safe system operation from the point of view of lead time.

The results of this section, while useful in their own right, are employed in Sections 4 and 5 for solving the LTC problem in the open- and closed-loop environments, respectively.



(a) L_1 (b) L_2 (c) L_3

Figure 3.2: Relative lead time, $\widehat{lt}$, as a function of relative workload, ρ_{max} (for $M = 10$, $\tau = 1$ min)

4 Open-Loop Control of Lead Time in Synchronous Exponential Lines

In this section, for both identical and non-identical producing machines, we first quantify the set of attainable lead times (feasible set) and then derive formulas for the release machine parameters that ensure the desired feasible lead time, while maximizing the throughput.

4.1 Identical producing machines

Proposition 4.1 *Under the assumptions of Proposition 3.1, the sets of feasible lead times, $\mathscr{F}_{\widehat{lt}}$ and $\mathscr{F}_{\widehat{lt}_\infty}$, are given, respectively, by*

$$\begin{aligned} \widehat{lt} &> 1 + (1-e)\frac{2M-1}{M}\frac{T_{down}}{\tau}, \\ \widehat{lt}_\infty &> 1 + 2(1-e)\frac{T_{down}}{\tau}. \end{aligned} \tag{4.1}$$

Proof: See the Appendix.

From these expressions, we observe that the lower bounds on $\widehat{lt}$ and $\widehat{lt}_\infty$ are decreasing functions of the producing machine efficiency and, for fixed e, increasing functions of the producing machine downtime in units of the cycle time. For instance, if $e = 0.8$ and $\frac{T_{down}}{\tau} = 10$, then $\widehat{lt} > 4.8$ (for $M = 10$) and $\widehat{lt}_\infty > 5$, no matter how low the release rate is.

Proposition 4.2 *Under the assumptions of Proposition 3.1, for any feasible desired lead time,*

$lt_d \in \mathscr{F}_{\widehat{lt}}$, the release rate is given by

$$\hat{e}_0^* = e\Big[1 - \frac{\mu + (2M-1)\mu_0}{M\mu\mu_0\tau(lt_d - 1) + \mu(1-e)}(1-e)\Big] \tag{4.2}$$

and, for this release rate,

$$\widehat{TP}^* = \frac{\hat{e}_0^*}{\tau}, \ \widehat{WIP}_0^* = \frac{\hat{e}_0^*}{\tau}\Big(\frac{\hat{e}_0^*}{\mu_0} + \frac{e}{\mu}\Big)\Big(\frac{1-e}{e-\hat{e}_0^*}\Big), \ \widehat{WIP}_i^* = \frac{2\hat{e}_0^* e}{\mu\tau}\Big(\frac{1-e}{e-\hat{e}_0^*}\Big), \ i = 1, 2, \ldots, M-1. \tag{4.3}$$

Proof: See the Appendix.

Note that the second term in the brackets of (4.2) is less than 1, as long as $lt_d \in \mathscr{F}_{\widehat{lt}}$.

This proposition leads to a solution of the open-loop LTC problem as follows:

- Since, as it is possible to show, $\frac{d\hat{e}_0^*}{d\mu_0} > 0$, $\widehat{TP}^*$ is maximized as $\mu_0 \to \infty$. In this case, the release rate that results in lt_d, while maximizing $\widehat{TP}^*$, becomes:

$$\hat{e}_0^*(\mu_0 = \infty) := \lim_{\mu_0 \to \infty} \hat{e}_0^* = e\Big[1 - \frac{(2M-1)(1-e)}{M(lt_d - 1)}\frac{T_{down}}{\tau}\Big]. \tag{4.4}$$

- Having $\mu_0 \to \infty$ with $\hat{e}_0^*$ being fixed, implies that $\lambda_0 \to \infty$ in such a manner that

$$\lim_{\substack{\lambda_0 \to \infty \\ \mu_0 \to \infty}} \frac{\lambda_0}{\mu_0} = \frac{1 - \hat{e}_0^*(\mu_0 = \infty)}{\hat{e}_0^*(\mu_0 = \infty)}. \tag{4.5}$$

In other words, both $T_{up,0}$ and $T_{down,0}$ tend to 0 and, thus, raw material is released continuously with the rate (4.4). In practice, this can be accomplished by releasing a part at the beginning of each cycle with probability

$$p = \hat{e}_0^*(\mu_0 = \infty). \tag{4.6}$$

This implies that the release machine can be viewed as obeying the Bernoulli reliability model with the probability of success given by (4.6). We refer to this type of release as *once-per-cycle*. In Subsection 4.3, it is generalized to a deterministic *once-per-hour* or *once-per-shift* release.

- In the limit as $M \to \infty$ and $lt_d \in \mathscr{F}_{\widehat{lt}_\infty}$, (4.2) becomes

$$\hat{e}_0^*(M = \infty) := \lim_{M\to\infty} \hat{e}_0^* = e\left[1 - \frac{2(1-e)}{lt_d - 1}\frac{T_{down}}{\tau}\right], \tag{4.7}$$

  which is independent of μ_0. Thus, for sufficiently large M, once-per-cycle release also can be implemented with

$$p = \hat{e}_0^*(M = \infty). \tag{4.8}$$

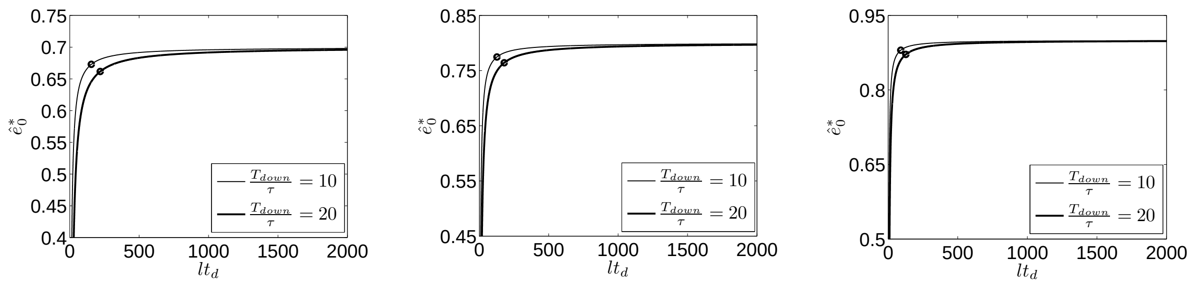
Summarizing the above arguments, we conclude that a solution of the open-loop LTC problem is provided by releasing a part into the raw material buffer b_0 once-per-cycle with probability (4.6) if M is relatively small (say, $M < 10$) and with probability (4.8) if $M \geqslant 10$.

The behavior of $\hat{e}_0^*(M = \infty)$ as a function of lt_d is illustrated in Figure 4.1 for various values of e and $\frac{T_{down}}{\tau}$, with black dots indicating $(\widehat{lt}_{knee}, \hat{e}_0^*(\widehat{lt}_{knee}))$. From this figure, we conclude:

- For $lt_d < \widehat{lt}_{knee}$, the optimal release rate $\hat{e}_0^*$ (and, therefore, $\widehat{TP}$) is a rapidly increasing function of lt_d.

- For $lt_d > \widehat{lt}_{knee}$, $\hat{e}_0^*$ is practically constant.

- Thus, *releasing the raw material beyond the knee is not only unnecessary* (since $\widehat{TP}$ practically does not grow), *but detrimental as well* (since $\widehat{WIP}$ grows almost linearly according to $\widehat{WIP} = \widehat{TP}(LT - M\tau)$).



(a) $e = 0.7$ (b) $e = 0.8$ (c) $e = 0.9$

Figure 4.1: Optimal release rate, $\hat{e}_0^*$, as a function of the desired relative lead time, lt_d, and machine parameters (for $M = 10$)

4.2 Non-identical producing machines

Proposition 4.3 *Under the assumptions of Proposition 3.3, the sets of feasible lead times, $\mathscr{F}_{\widehat{lt}}$, $\mathscr{F}_{\widehat{\overline{lt}}}$, and $\mathscr{F}_{\widehat{\overline{lt}}_\infty}$, are given, respectively, by*

$$\begin{aligned} \widehat{lt} &> 1 + \frac{1}{M\tau}\Big(\sum_{i=1}^{M}\frac{1-e_i}{\mu_i} + \sum_{i=1}^{M-1}\frac{e_i(1-e_{i+1})}{\mu_i e_{i+1}}\Big), \\ \widehat{\overline{lt}} &> 1 + (1-e_{min})\frac{2M-1}{M\tau\mu_{min}}\frac{e_{max}}{e_{min}}, \\ \widehat{\overline{lt}}_\infty &> 1 + \frac{2(1-e_{min})}{\tau\mu_{min}}\frac{e_{max}}{e_{min}}. \end{aligned} \tag{4.9}$$

where, as before, $e_{min} = \min_{1\leqslant i\leqslant M} e_i$, $e_{max} = \max_{1\leqslant i\leqslant M} e_i$, and $\mu_{min} = \min_{1\leqslant i\leqslant M}\mu_i$.

Proof: See the Appendix.

Proposition 4.4 *Under the assumptions of Proposition 3.3, for any feasible desired lead time, $lt_d \in \mathscr{F}_{\widehat{lt}}$, the release rate $\hat{e}_0^*$ that ensures this lead time is the unique real root less than $\min_{1\leqslant i\leqslant M} e_i$ of the following M-th order polynomial equation:*

$$(LT_d - M\tau)\prod_{i=0}^{M-1}(e_{i+1}-e_0) - (1-e_1)\Big(\frac{e_0}{\mu_0}+\frac{e_1}{\mu_1}\Big)\prod_{i=1}^{M-1}(e_{i+1}-e_0) - \sum_{i=1}^{M-1}\Big((1-e_{i+1})\Big(\frac{e_i}{\mu_i}+\frac{e_{i+1}}{\mu_{i+1}}\Big)\prod_{j=0,j\neq i}^{M-1}(e_{j+1}-e_0)\Big) = 0 \tag{4.10}$$

and

$$\widehat{TP}^* = \frac{\hat{e}_0^*}{\tau},\ \widehat{WIP}_0^* = \frac{\hat{e}_0^*}{\tau}\Big(\frac{\hat{e}_0^*}{\mu_0}+\frac{e_1}{\mu_1}\Big)\Big(\frac{1-e_1}{e_1-\hat{e}_0^*}\Big),\ \widehat{WIP}_i^* = \frac{\hat{e}_0^*}{\tau}\Big(\frac{e_i}{\mu_i}+\frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-\hat{e}_0^*}\Big),\ i = 1,2,\ldots,M-1. \tag{4.11}$$

Proof: See the Appendix.

For instance, if $M = 2$, Equation (4.10) takes the form

$$\begin{aligned} &\Big(LT_d - 2\tau + \frac{1-e_1}{\mu_0}\Big)e_0^2 - \Big[(LT_d - 2\tau)(e_1+e_2) + (1-e_1)\Big(\frac{e_2}{\mu_0}-\frac{e_1}{\mu_1}\Big) - (1-e_2)\Big(\frac{e_1}{\mu_1}+\frac{e_2}{\mu_2}\Big)\Big]e_0 \\ &+(LT_d - 2\tau)e_1e_2 - (1-e_1)\frac{e_1e_2}{\mu_1} - (1-e_2)\Big(\frac{e_1}{\mu_1}+\frac{e_2}{\mu_2}\Big)e_1 = 0, \end{aligned} \tag{4.12}$$

and, thus, the release rate is given by

$$\hat{e}_0^* = \frac{(LT_d - 2\tau)(e_1 + e_2) + (1 - e_1)\left(\frac{e_2}{\mu_0} - \frac{e_1}{\mu_1}\right) - (1 - e_2)\left(\frac{e_1}{\mu_1} + \frac{e_2}{\mu_2}\right) - \sqrt{\Delta}}{2\left(LT_d - 2\tau + \frac{1-e_1}{\mu_0}\right)}, \tag{4.13}$$

where $\Delta = \left[(LT_d - 2\tau)(e_2 - e_1) + (1 - e_1)\left(\frac{e_2}{\mu_0} - \frac{e_1}{\mu_1}\right) - (1 - e_2)\left(\frac{e_1}{\mu_1} + \frac{e_2}{\mu_2}\right)\right]^2 + 4(LT_d - 2\tau)(1 - e_1)(e_2 - e_1)\frac{e_1}{\mu_1} + 4\frac{e_1(1-e_1)}{\mu_0}\left[(1 - e_1)\frac{e_2}{\mu_1} + (1 - e_2)\left(\frac{e_1}{\mu_1} + \frac{e_2}{\mu_2}\right)\right]$.

For $M > 2$, solving equation (4.10) might be too complex for practical applications. Therefore, using the upper bounds $\overline{\overline{\widehat{lt}}}$ and $\overline{\overline{\widehat{lt}}}_\infty$, we provide below lower bounds on $\hat{e}_0^*$, which could be useful in applications.

Proposition 4.5 *Let $\bar{\hat{e}}_0^*$ and $\bar{\hat{e}}_{0,\infty}^*$ be the release rates that solve the open-loop LTC problem for $\overline{\overline{\widehat{lt}}}$ and $\overline{\overline{\widehat{lt}}}_\infty$ with $lt_d \in \{\mathscr{F}_{\widehat{lt}} \cap \mathscr{F}_{\overline{\overline{\widehat{lt}}}} \cap \mathscr{F}_{\overline{\overline{\widehat{lt}}}_\infty}\}$. Then,*

$$\bar{\hat{e}}_0^* = e_{min}\left[1 - \frac{\mu_{min} + (2M-1)\mu_0\frac{e_{max}}{e_{min}}}{M\mu_{min}\mu_0\tau(lt_d - 1) + \mu_{min}(1 - e_{min})}(1 - e_{min})\right], \tag{4.14}$$

$$\bar{\hat{e}}_{0,\infty}^* = e_{min}\left[1 - \frac{2(1 - e_{min})\frac{e_{max}}{e_{min}}}{\tau\mu_{min}(lt_d - 1)}\right], \tag{4.15}$$

and, if $\mu_0 \geqslant \mu_{min}$,

$$\hat{e}_0^* \geqslant \bar{\hat{e}}_0^* \geqslant \bar{\hat{e}}_{0,\infty}^*. \tag{4.16}$$

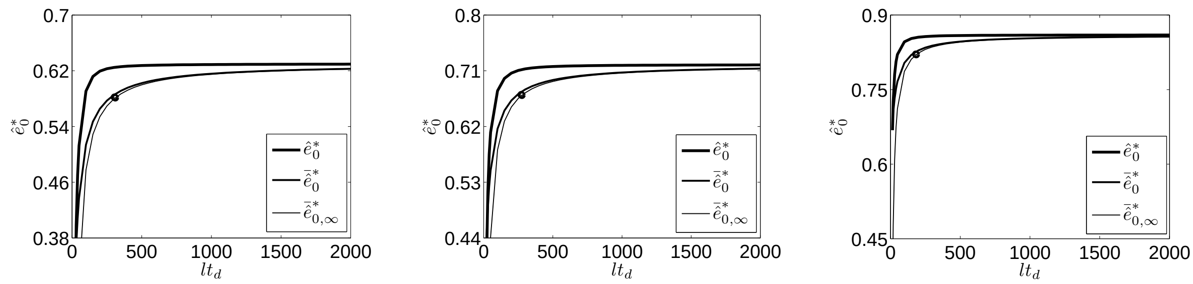
Proof: See the Appendix.

Similar to the identical machine case, the solution of the open-loop LTC problem for non-identical machines can be implemented by releasing raw material once-per-cycle with probability $\bar{\hat{e}}_0^*$ given by (4.14) with $\mu_0 = \infty$, i.e.,

$$p = e_{min}\left[1 - \frac{(2M-1)\frac{e_{max}}{e_{min}}}{M\mu_{min}\tau(lt_d - 1)}(1 - e_{min})\right], \tag{4.17}$$

or with $p = \bar{\hat{e}}_{0,\infty}^*$ given by (4.15).

The behavior of $\hat{e}_0^*$, $\bar{\hat{e}}_0^*$, and $\bar{\hat{e}}_{0,\infty}^*$ is illustrated in Figure 4.2 (where black dots indicate the knee) as a function of lt_d for the three lines given in (3.22). From this figure, we conclude that, similar to the identical machine case, raw material release with rates beyond the knee is not only unnecessary, but detrimental as well (since TP is practically constant and $\widehat{WIP} = \widehat{TP}(LT_d - M\tau)$).



(a) L_1 (b) L_2 (c) L_3

Figure 4.2: Release rates, $\hat{e}_0^*$, $\bar{\hat{e}}_0^*$, and $\bar{\hat{e}}_{0,\infty}^*$, as a function of the desired relative lead time, lt_d (for $M = 10$)

4.3 Deterministic once-per-hour or once-per-shift release

In practice, random, once-per-cycle, raw material release may be inconvenient. In such situations, the results of Subsections 4.1 and 4.2 can be used to define strategies for deterministic release per a fixed interval of time, say, once-per-hour or once-per-shift. This is carried out below.

Let $\hat{e}_0^*(lt_d)$ be a once-per-cycle release rate calculated using either (4.2), (4.4), (4.7), or (4.10). Then, the deterministic hourly release, $\hat{E}_H^*$ (parts/hour), can be defined as:

$$\hat{E}_H^* = \lfloor H\hat{e}_0^*(lt_d) \rfloor, \tag{4.18}$$

where $\lfloor x \rfloor$ is the “floor” operator, which denotes the largest integer not greater than x, and H is the number of cycle times in an hour, i.e.,

$$H = \frac{60}{\tau}. \tag{4.19}$$

Releasing each hour the amount of raw material defined by (4.18), leads to the following inequality:

$$\widehat{LT}(\hat{E}_H^*) < \widehat{LT}(\hat{e}_0^*) + 60, \tag{4.20}$$

where $\widehat{LT}(\hat{e}_0^*)$ and $\widehat{LT}(\hat{E}_H^*)$ are the lead times for per-cycle and per-hour release, respectively.

When a solution of (4.10) is not available, Equations (4.14) and (4.15) can be used to evaluate the lower bounds of $\hat{e}_0^*$ and then the hourly release calculated as

$$\bar{\hat{E}}_H^* = \left\lfloor H\bar{\hat{e}}_0^*(lt_d) \right\rfloor \text{ or } \bar{\hat{E}}_{H,\infty}^* = \left\lfloor H\bar{\hat{e}}_{0,\infty}^*(lt_d) \right\rfloor. \tag{4.21}$$

In this case,

$$\widehat{LT}(\bar{\hat{E}}_H^*) < \widehat{LT}(\bar{\hat{e}}_0^*) + 60 \tag{4.22}$$

and

$$\widehat{LT}(\bar{\hat{E}}_{H,\infty}^*) < \widehat{LT}(\bar{\hat{e}}_{0,\infty}^*) + 60. \tag{4.23}$$

The tightness of bound (4.20) has been evaluated under hourly release for various τ and LT_d by simulating three synchronous exponential lines with ten identical machines and with parameters

$$L_1 : \ e = 0.9, \ T_{down} = 70; \quad L_2 : \ e = 0.9, \ T_{down} = 7; \quad L_3 : \ e = 0.9, \ T_{down} = 0.7. \tag{4.24}$$

The lead time and throughput of each line has been evaluated based on the same simulation procedure as in Subsection 3.1. The lt_d for these simulations has been selected so that, on one hand, it is in the admissible domain (defined by (4.1)) and, on the other hand, the system parameters are in the sets (3.2), which render sufficiently high accuracy of $\widehat{lt}$ (defined by (3.1)). Based on lt_d, thus selected, $\hat{e}_0^*$ and $\hat{E}_H^*$ have been evaluated using (4.2) and (4.18), respectively. For each system considered, we ran the simulations with once-per-cycle and once-per-hour release and evaluated the resulting lead times, lt_c and lt_H, and throughputs, TP_c and TP_H, where the subscripts "c" and "H" stand for cycle and hour, respectively. Based on these measurements, we quantified changes in lt and losses in TP by

$$\begin{aligned} lt_{change} &= \frac{lt_H - lt_c}{lt_c} \times 100\%, \\ TP_{loss} &= \frac{TP_c - TP_H}{TP_c} \times 100\%. \end{aligned} \tag{4.25}$$

The results are shown in Tables 4.1 and 4.2 for $\tau = 0.5$ min and $\tau = 5$ min, respectively. From these data, we conclude:

- When $\tau = 0.5$ min, TP_{loss} is insignificant, while lt_{change} may be large. The significant values of lt_{change} are due to the "floor" operator in (4.18) and the waiting time in the raw material buffer under hourly release. More specifically, for small lt_d, the latter is more important and $lt_c < lt_H$; for large lt_d, the significance of these two causes are reversed, and $lt_c > lt_H$.

- When $\tau = 5$ min, TP_{loss} may be quite significant. The reason for this is that, for large τ, the amount of material released per-hour amounts to just a few parts, even if lt_d is large. To combat this problem, a release for a longer interval of time (e.g., once-per-shift, rather than once-per-hour) may be considered. In the case of an eight-hour shift, the release becomes

$$\hat{E}_S^* = \left\lfloor \frac{480}{\tau} \hat{e}_0^*(lt_d) \right\rfloor \tag{4.26}$$

where, as before, $\hat{e}_0^*(lt_d)$ is the release rate per-cycle that ensures lt_d. The results for this release are shown in Table 4.3. As one can see, these data are quite similar to those of Table 4.1. Based on this and some additional experiments, we conclude that the release interval (RI), which leads to practically no changes in the throughput (as compared with once-per-cycle release), can be defined as $RI \geqslant 50\tau$.

5 Closed-Loop Control of Lead Time in Synchronous Exponential Lines

5.1 Scenario

The previous section provided methods for calculating raw material release rates that ensure the desired lead time, given that the parameters of the machines are known precisely. In practice, however, this is seldom the case – the real values of machine parameters (e.g, their efficiencies or up- and downtimes) are often unknown; only their nominal values are available. In this situation, the above methods may result in lead times dramatically different from the expected ones. Indeed, if, for example, the real machine efficiency, e_{real}, is lower than the nominal one, e_{nom}, and the desired lead time, lt_d, is sufficiently large, it may happen that

$$\hat{e}_0^*(lt_d) > \min_{1 \leqslant i \leqslant M} e_{real,i}, \tag{5.1}$$

Table 4.1: Lead time, $LT(\mathscr{R}^*)$, under once-per-hour release for serial lines with identical synchronous exponential machines ($\tau = 0.5$ min)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_c	lt_H	lt_{change} (%)	TP_c	TP_H	TP_{loss} (%)
90	0.6212	74	85.84	68.41	−20.30	1.2426	1.2333	0.74
120	0.6907	82	118.57	93.10	−21.48	1.3809	1.3667	1.03
300	0.8161	97	308.84	235.99	−23.59	1.6318	1.6167	0.93
1500	0.8832	105	1543.21	871.26	−43.54	1.7665	1.7500	0.93

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_c	lt_H	lt_{change} (%)	TP_c	TP_H	TP_{loss} (%)
12	0.6738	80	11.75	13.53	15.15	1.3479	1.3333	1.08
16	0.7336	88	16.09	17.66	9.78	1.4671	1.4667	0.03
40	0.8356	100	41.27	38.28	−7.24	1.6713	1.6666	0.28
200	0.8873	106	202.43	137.67	−31.99	1.7745	1.7667	0.44

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_c	lt_H	lt_{change} (%)	TP_c	TP_H	TP_{loss} (%)
4.5	0.8283	97	4.61	9.42	104.58	1.656	1.6500	0.38
6	0.8497	101	6.14	10.30	67.65	1.6993	1.6833	0.94
15	0.8820	105	15.31	15.36	0.37	1.7638	1.7500	0.78
75	0.8966	107	73.90	32.55	−55.96	1.7932	1.7833	0.55

resulting in an arbitrarily large lead time.

To prevent this situation, feedback control can be used to throttle the raw material release if the work-in-process in the systems exceeds a certain limit. A number of such control strategies can be proposed. Here, we investigate the one which is simple enough for factory floor implementations. Specifically, we consider hourly release based on the real-time total work-in-process, WIP_{total}: if it is below a threshold defined by the nominal WIP, the raw material is released; otherwise it is not. In Subsection 5.2 below we introduce this control law and in Subsection 5.3 investigate its performance using simulations.

5.2 Control Law

Consider a synchronous exponential serial line defined by the nominal breakdown and repair rates λ_i and μ_i, $i = 1, 2, \ldots, M$, respectively. Let LT_d be the desired lead time. Based on this information, calculate:

Table 4.2: Lead time, $LT(\mathscr{R}^*)$, under once-per-hour release for serial lines with identical synchronous exponential machines ($\tau = 5$ min)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_c	lt_H	lt_{change} (%)	TP_c	TP_H	TP_{loss} (%)
12	0.6738	8	11.75	9.43	−19.74	0.1348	0.1333	1.10
16	0.7336	8	16.02	9.43	−41.13	0.1466	0.1333	9.06
40	0.8356	10	41.15	33.65	−18.23	0.1671	0.1667	0.27
200	0.8873	10	199.17	33.65	−83.10	0.1774	0.1667	6.08

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_c	lt_H	lt_{change} (%)	TP_c	TP_H	TP_{loss} (%)
4.5	0.8283	9	4.62	2.83	−38.82	0.1657	0.1500	9.47
6	0.8497	10	6.14	4.73	−22.95	0.1699	0.1667	1.90
15	0.8820	10	15.47	4.73	−69.44	0.1764	0.1667	5.54
75	0.8966	10	76.19	4.72	−93.80	0.1793	0.1667	7.06

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_c	lt_H	lt_{change} (%)	TP_c	TP_H	TP_{loss} (%)
1.5	0.8497	10	1.51	1.88	24.20	0.1699	0.1667	1.92
2	0.8748	10	2.01	1.88	−6.60	0.1749	0.1667	4.73
5	0.8937	10	5.00	1.88	−62.35	0.1787	0.1667	6.73
25	0.8990	10	24.45	1.88	−92.31	0.1798	0.1667	7.30

- Per-cycle release rate, $\hat{e}_0^*$, using either (4.6) or (4.10).
- Per-hour release rate, $\hat{E}_H^*$, using (4.18) or per-shift release rate, $\hat{E}_S^*$, using (4.26).
- The nominal total work-in-process in the system. As it follows from (3.1) and (3.16),

$$\widehat{WIP}_{total} = \frac{\hat{e}_0^*}{\tau}(LT_d - M\tau). \tag{5.2}$$

Using these data, introduce the following control law for raw material release:

$$E(s+1) = \begin{cases} \hat{E}^*, & \text{if } WIP_{total}(s) \leqslant \widehat{WIP}_{total}, \\ 0, & \text{otherwise,} \end{cases} \tag{5.3}$$

$$s = 0, 1, \ldots,$$

Table 4.3: Lead time, $LT(\mathscr{R}^*)$, under once-per-shift release for serial lines with identical synchronous exponential machines ($\tau = 5$ min)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_c	lt_S	lt_{change} (%)	TP_c	TP_S	TP_{loss} (%)
12	0.6738	64	11.75	12.57	7.02	0.1348	0.1333	1.10
16	0.7336	70	16.02	16.32	1.89	0.1466	0.1458	0.53
40	0.8356	80	41.15	37.26	−9.46	0.1671	0.1667	0.27
200	0.8873	85	199.17	154.08	−22.64	0.1774	0.1771	0.21

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_c	lt_S	lt_{change} (%)	TP_c	TP_S	TP_{loss} (%)
4.5	0.8283	79	4.62	8.19	77.27	0.1657	0.1646	0.67
6	0.8497	81	6.14	9.28	51.30	0.1699	0.1687	0.68
15	0.8820	84	15.47	14.23	−8.03	0.1764	0.1750	0.82
75	0.8966	86	76.19	57.31	−24.78	0.1793	0.1792	0.09

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_c	lt_S	lt_{change} (%)	TP_c	TP_S	TP_{loss} (%)
1.5	0.8497	81	1.51	6.05	299.53	0.1699	0.1687	0.69
2	0.8748	83	2.01	6.30	212.87	0.1749	0.1729	1.16
5	0.8937	85	5.00	7.18	43.72	0.1787	0.1771	0.90
25	0.8990	86	24.45	10.77	−55.97	0.1798	0.1791	0.38

where $s = 0, 1, \ldots,$ is the index of hour/shift; $E(s + 1)$ is the raw material release at the beginning of hour/shift $s + 1$; $\hat{E}^* = \hat{E}_H^*$ for hourly release and $\hat{E}^* = \hat{E}_S^*$ for release per shift; and $WIP_{total}(s)$ is the real-time total work-in-process in the system at the end of the hour/shift s.

Clearly, the "sensor measurement" in this control law is $WIP_{total}(s)$, $s = 0, 1, \ldots$. In some production system this information is readily available from manufacturing monitoring systems; in other it is not. In the latter case, the following simple calculation can be used to evaluate $WIP_{total}(s)$:

$$WIP_{total}(s + 1) = WIP_{total}(s) + E(s + 1) - N(s + 1), \ s = 0, 1, \ldots, \tag{5.4}$$

where $N(s + 1)$ is the number of parts produced during the hour/shift $s + 1$. Using (5.4) the only input to the control law (5.3) is the initial value of WIP, i.e., $WIP_{total}(0)$.

5.3 Performance evaluation

To evaluate the performance of feedback law (5.3), we use the three exponential lines (4.24) as nominal ones and form a real one for each of them. The real lines are formed by increasing or decreasing machine up- and downtimes randomly and equiprobably within ±50% of their nominal values. The resulting lines are as follows:

$$
\begin{aligned}
L_1 :\ & \boldsymbol{e} = [0.93, 0.89, 0.94, 0.91, 0.86, 0.92, 0.84, 0.93, 0.93, 0.83],\\
& \boldsymbol{T_{down}} = [45.46, 83.00, 51.47, 35.40, 97.05, 81.68, 98.71, 61.90, 55.10, 79.16],\\
L_2 :\ & \boldsymbol{e} = [0.83, 0.94, 0.91, 0.90, 0.88, 0.91, 0.90, 0.95, 0.90, 0.84],\\
& \boldsymbol{T_{down}} = [7.66, 4.24, 8.62, 9.72, 10.12, 5.06, 6.43, 4.27, 6.55, 7.45],\\
L_3 :\ & \boldsymbol{e} = [0.94, 0.89, 0.89, 0.91, 0.92, 0.95, 0.91, 0.91, 0.93, 0.79],\\
& \boldsymbol{T_{down}} = [0.51, 0.84, 0.44, 0.74, 0.51, 0.36, 0.68, 0.83, 0.38, 1.00].
\end{aligned}
\tag{5.5}
$$

We simulated these lines with and without feedback control (5.3) for $\tau = 0.5$ min with hourly release and for $\tau = 5$ min with release per shift. The simulations have been carried out using the procedure described in Subsection 3.1. Based on these simulations, the lead times in open- and closed-loop cases (denoted as lt_{OL} and lt_{CL}) have been evaluated. The results are shown in Tables 5.1 and 5.2. From these data we conclude that, in all cases considered, closed-loop raw material release maintains the lead time close to the desired one, whereas the open-loop release results in a substantially longer lead time, and becomes unbounded for large lt_d. Thus, the proposed control law (5.3) is indeed effective in constraining production lead time in real systems.

6 Extensions

In this section, we present initial results on extending methods described above to asynchronous exponential and synchronous non-exponential lines. Although more research in both of these directions is necessary, the results obtained show that extensions to these larger classes of production systems are indeed possible.

Table 5.1: Lead time, LT, under control law (5.3) ($\tau = 0.5$ min, once-per-hour release)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_{OL}	lt_{CL}
150	0.7324	87	154.63	110.50
300	0.8161	97	521.06	255.05
600	0.8580	102	∞	581.96
1500	0.8832	105	∞	1575.48
3000	0.8916	106	∞	3215.71

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_{OL}	lt_{CL}
20	0.7683	92	27.47	19.67
40	0.8356	100	∞	38.45
80	0.8682	104	∞	81.51
200	0.8873	106	∞	213.25
400	0.8937	107	∞	431.84

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_H^*$	lt_{OL}	lt_{CL}
7	0.8581	102	∞	11.04
14	0.8806	105	∞	16.24
28	0.8907	106	∞	32.23
70	0.8963	107	∞	80.31
140	0.8982	107	∞	160.29

6.1 Asynchronous exponential lines

The approach of this subsection is based on introducing an *auxiliary synchronous* exponential line and showing, by simulations, that the lead time of the auxiliary line provides an upper bound for the lead time of the *original asynchronous* one.

Consider an asynchronous exponential line defined by assumptions (i)-(v). Note that according to assumptions (ii) and (iii), each producing machine cycle time is τ_i, $i = 1, 2, \ldots, M$, its breakdown and repair rates are λ_i and μ_i, respectively, and its throughput in isolation, $TP_{isol,i}$, is given by (2.2). Without loss of generality, assume that the cycle time of the release machine is defined by

$$\tau_0 = \min_{1\leqslant i\leqslant M} \tau_i \tag{6.1}$$

and, to obtain meaningful results, the breakdown and repair rates of the release machine, λ_0 and

Table 5.2: Lead time, LT, under control law (5.3) ($\tau = 5$ min, once-per-shift release)

(a) L_1

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_{OL}	lt_{CL}
20	0.7683	73	25.98	18.41
40	0.8356	80	∞	38.26
80	0.8682	83	∞	81.70
200	0.8873	85	∞	213.34
400	0.8937	85	∞	431.34

(b) L_2

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_{OL}	lt_{CL}
7	0.8581	82	∞	9.99
14	0.8806	84	∞	15.23
28	0.8907	85	∞	30.42
70	0.8963	86	∞	76.29
140	0.8982	86	∞	152.60

(c) L_3

lt_d	$\hat{e}_0^*$	$\hat{E}_S^*$	lt_{OL}	lt_{CL}
4	0.8916	85	∞	7.87
8	0.8964	86	∞	10.19
16	0.8983	86	∞	18.76
40	0.8994	86	∞	46.16
80	0.8997	86	∞	91.04

μ_0, are selected so that

$$TP_{isol,0} < \min_{1 \leqslant i \leqslant M} TP_{isol,i}. \tag{6.2}$$

Thus, the relative load factor, defined as

$$\rho_{async} := \frac{TP_{isol,0}}{\min\limits_{1 \leqslant i \leqslant M} TP_{isol,i}}, \tag{6.3}$$

is less than 1.

Along with this asynchronous line, consider an auxiliary synchronous line with the producing machines defined as follows:

$$\check{\tau} = \tau_0, \tag{6.4}$$

$$\check{\mu}_i = \mu_i, \; i = 1, 2, \ldots, M, \tag{6.5}$$

$$\check{TP}_{isol,i} = TP_{isol,i},\ i = 1, 2, \dots, M. \tag{6.6}$$

From (6.5) and (6.6) it follows that

$$\check{e}_i = \frac{\check{\tau}}{\tau_i} e_i, \tag{6.7}$$

$$\check{\lambda}_i = \frac{\mu_i}{\check{e}_i}(1 - \check{e}_i). \tag{6.8}$$

The release machine of the auxiliary line remains the same as in the asynchronous one, i.e., defined by (6.1) and (6.2). Therefore, the relative load factor for the auxiliary line, ρ_{sync}, is the same as for the asynchronous one, i.e., given by (6.3). For the sake of brevity, we omit the subscript of both load factors and denote them as ρ.

Let $\mathcal{LT}_{async}$ and $\mathcal{LT}_{sync}$ denote the lead times of the original asynchronous and the auxiliary synchronous lines. We address two problems concerning to these two measures: the first one is related to a bound between them and the second to the tightness of this bound. Both problems are analyzed using simulations on the lines in question. To investigate the first one, we formed 1000 asynchronous lines with parameters selected randomly and equiprobably from the following sets:

$$\begin{aligned}
&M \in [3, 10],\ \tau_i \in [0.8\text{min}, 1.2\text{min}],\ e_i \in [0.7, 0.99],\ i = 1, 2, \dots, M,\\
&TP_{isol,0} \in [0.7 \min_{1\leqslant i\leqslant M} TP_{isol,i},\ 0.99 \min_{1\leqslant i\leqslant M} TP_{isol,i}],\\
&T_{down,i} \in [10\text{min}, 100\text{min}],\ i = 0, 1, \dots, M.
\end{aligned} \tag{6.9}$$

For each of these asynchronous lines, we form an auxiliary synchronous one according to (6.4)-(6.8) and simulate 1000 pairs of lines using the procedure described in Section 3. As a result, we obtain:

Numerical Fact 6.1 *For all* 1000 *pairs of lines analyzed,* $\mathcal{LT}_{async} < \mathcal{LT}_{sync}$, *i.e., the lead time of the auxiliary synchronous line is an upper bound of the lead time of the original asynchronous one.*

The tightness of this bound is analyzed. Let $\mathcal{LT}_{async}$ and $\mathcal{LT}_{sync}$ denote the lead times of the original asynchronous and the auxiliary synchronous lines. We quantify the difference between them by

$$\Delta_{LT} = \frac{LT_{sync} - LT_{async}}{LT_{async}} \times 100\%. \tag{6.10}$$

For the 1000 pairs of lines analyzed, the smallest and largest values of Δ_{LT} were 2.73% and 169.90%, respectively, and the average value was 52.21%.

To further investigate the behavior of Δ_{LT} as a function of ρ, we generate 100 asynchronous lines and form their corresponding auxiliary synchronous lines according to (6.4)-(6.8). The asynchronous lines are generated based on the ten machine Line L_1 given in (5.5). The cycle time, efficiency, and downtime of the producing machines are selected randomly and equiprobably from the following sets

$$\tau_i \in [0.8\text{min}, 1.2\text{min}],\ e_i \in [0.9e'_i, \min(1.1e'_i, 1)],\ T_{down,i} \in [0.9T'_{down,i}, 1.1T'_{down,i}],\ i = 1, 2, \ldots, M, \tag{6.11}$$

where e'_i and $T'_{down,i}$ are respectively machine efficiency and downtime of Line L_1 in (5.5); the load factor ρ is selected from set

$$\rho \in \{0.7, 0.8, 0.9, 0.99\}. \tag{6.12}$$

Simulating these lines and the corresponding auxiliary lines, we obtained the results illustrated in Figure 6.1. From these 100 lines analyzed, we conclude that Δ_{LT} is practically independent of ρ. Thus,

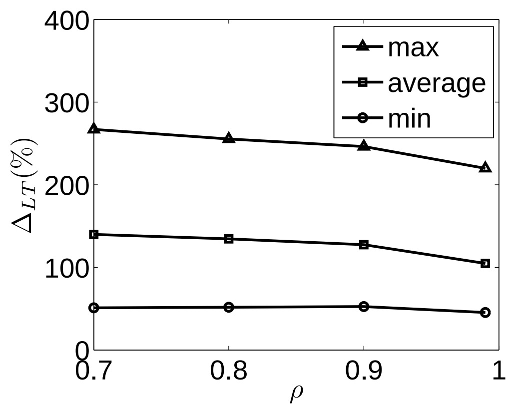


Figure 6.1: Tightness of lead time upper bound

6.2 Non-exponential lines

In this subsection we study the lead time in synchronous serial lines with identical machines having non-exponential reliability models and show, by simulations, that LT is a monotonically increasing function of the up- and downtime coefficient of variation (CV) for $CV \in [0, 1]$. In addition, as it turns out, this function is practically the same for all reliability models considered; this leads to an empirical formula for LT as a function of CV. Since the extent of this simulation study is quite limited, additional research in this direction is desirable.

The systems considered here are ten-machine lines with $\tau = 1$ min, $T_{down} = 5$ min, and the other parameters selected as combination from the following sets:

$$\begin{aligned}
&e \in \{0.7, 0.8, 0.9\},\ \rho \in \{0.7, 0.8, 0.9, 0.99\},\\
&\text{Reliability model} \in \{\text{Weibull, gamma, log-normal}\},\\
&CV \in \{0.01, 0.1, 0.25, 0.5, 0.75, 1\},
\end{aligned} \tag{6.13}$$

where ρ, as before, is defined in (3.3). The reason for selecting $CV \leqslant 1$ is in the following: An empirical study, reported in [37], showed that most of manufacturing equipment in the automotive industry has the CV of up- and downtime less than 1. Analytical studies of [.] and [.] proved that if the breakdown and repair rates are increasing functions of time (implying, for example, that the longer machine is up, the larger is the probability that it breaks down in the ensuing infinitesimal time interval), the respective CV's are again less than 1. Thus, the assumption $CV \leqslant 1$ is supported by both practical and theoretical considerations.

Selecting the parameters from (6.13), we formed and simulated (using the procedure described in Section 3) 216 serial lines with identical producing machines and the release machine specified by ρ. The values of lt obtained by simulations are shown in Table 6.1 by broken lines. From these data, we conclude:

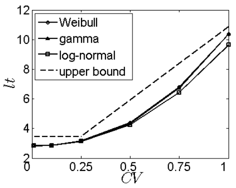
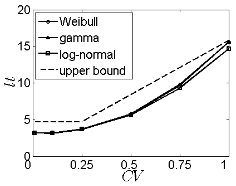
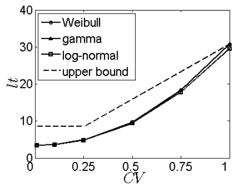
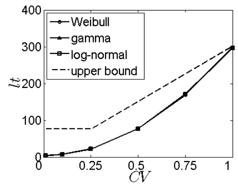
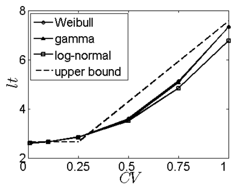
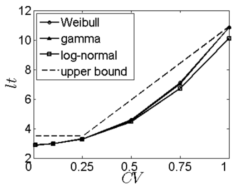
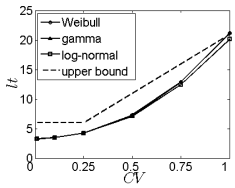
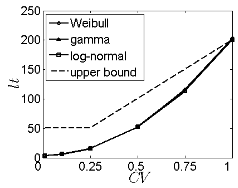
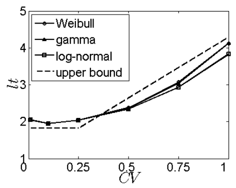
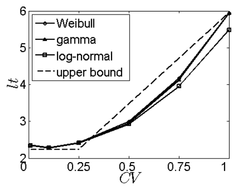
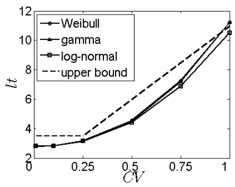
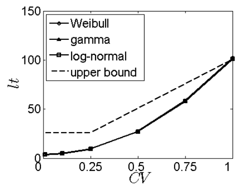
- lt is practically an increasing function of CV for $CV \leqslant 1$. In other words, lt_{exp} is an upper bound of lt for any $CV \leqslant 1$.
- The difference of lt for different reliability models decreases as ρ increases.

In addition, the curves of Table 6.1 indicate that lt can be upper bounded by a piecewise linear function shown in Table 6.1 by solid lines. An analytical representation of these lines is as follows: Since Weibull and gamma distributions with $CV = 1$ coincide with the exponential one, the increasing part of this upper bound connects the point $(0, 0)$ with the point $(1, \widehat{lt}_{exp})$, where $\widehat{lt}_{exp}$ can be calculated using (3.5). The constant part of the upper bound equals the value of the increasing part at $CV = 0.25$. Thus, the empirical upper bound of lt can be given as follows:

$$lt \leqslant \begin{cases} 0.25(\widehat{lt}_{exp} - 1) + 1, & \text{for } 0 < CV \leqslant 0.25, \\ (\widehat{lt}_{exp} - 1)CV + 1, & \text{for } 0.25 < CV \leqslant 1. \end{cases} \tag{6.14}$$

We hypothesize that, under some mild conditions (e.g., the up- and downtime obey a unimodal probability density function), bound (6.14) is applicable to any model of machine reliability. A verification of this hypothesis and its extension to systems with non-identical machines are topics of future work.

Table 6.1: Lead time for synchronous non-exponential lines with identical machines ($M = 10$)

e	ρ			
	0.7	0.8	0.9	0.99
0.7				
0.8				
0.9				

7 Conclusions and Future Work

This paper provides a method for calculating release rates of raw material leading to the desired lead time in Bernoulli and synchronous exponential lines. Although the method is approximate, taking into account "fuzziness" of machine parameter information available of the factory floor, the accuracy of the method is sufficient for practical applications. This method may be particularly useful for small and mid-size enterprises, where neither finite buffers nor feedback tools (e.g., kanban or CONWIP) are available to limit the inventories.

Future work in this area will be centered on extending the results to asynchronous exponential lines, non-exponential lines and, most importantly, to re-entrant lines, where large and unpredictable lead times often mar the performance.

Appendix

The analysis of $\widehat{LT}$ for synchronous exponential lines in this paper is based on the recursive aggregation procedure described in [3]. For serial lines with $M+1$ synchronous exponential machines defined by $(\lambda_0, \mu_0), (\lambda_1, \mu_1), \ldots, (\lambda_M, \mu_M)$ and M buffers with capacity $N_0, N_1, \ldots, N_{M-1}$, the steady state of this procedure, λ_i^f, μ_i^f, $i = 1, 2, \ldots, M$, and λ_i^b, μ_i^b, $i = 0, 1, \ldots, M-1$, is the unique solution of the following system of transcendental equations:

$$\begin{aligned}
\mu_i^f &= \mu_i - \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, N_{i-1}),\ 1 \leqslant i \leqslant M,\\
\lambda_i^f &= \lambda_i + \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, N_{i-1}),\ 1 \leqslant i \leqslant M,\\
\mu_i^b &= \mu_i - \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, N_i),\ 0 \leqslant i \leqslant M-1,\\
\lambda_i^b &= \lambda_i + \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, N_i),\ 0 \leqslant i \leqslant M-1,
\end{aligned} \tag{A.1}$$

with the boundary conditions $\lambda_0^f = \lambda_0$, $\mu_0^f = \mu_0$ and $\lambda_M^b = \lambda_M$, $\mu_M^b = \mu_M$ and

$$Q(x_1, y_1, x_2, y_2, N) = \begin{cases} \frac{(1-e_1)(1-\phi)}{1-\phi e^{-\beta N}}, & \text{if } \frac{x_1}{y_1} \neq \frac{x_2}{y_2},\\ \frac{x_1(x_1+x_2)(y_1+y_2)}{(x_1+y_1)[(x_1+x_2)(y_1+y_2)+x_2 y_1(x_1+x_2+y_1+y_2)N]}, & \text{if } \frac{x_1}{y_1} = \frac{x_2}{y_2}, \end{cases} \tag{A.2}$$

where

$$
\begin{aligned}
e_i &= \frac{y_i}{x_i + y_i}, \ i = 1, 2, \\
\phi &= \frac{e_1(1-e_2)}{e_2(1-e_1)}, \\
\beta &= \frac{(x_1 + x_2 + y_1 + y_2)(x_1 y_2 - x_2 y_1)}{(x_1 + x_2)(y_1 + y_2)}.
\end{aligned} \tag{A.3}
$$

The proofs of Propositions 3.1 and 3.3 are based on (A.1)-(A.3). Therefore, below we evaluate (A.2) and the solutions of (A.1) for $N_i = \infty$ (Lemmas A.1 and A.2, respectively) and then prove the above mentioned theorems.

Lemma A.1 *Function $Q(x_1, y_1, x_2, y_2, N)$, defined by (A.2) and (A.3), has the following limit:*

$$
\lim_{N\to\infty} Q(x_1, y_1, x_2, y_2, N) = \begin{cases} 0, & \text{if } \frac{x_1}{y_1} \leqslant \frac{x_2}{y_2}, \\ 1 - \frac{e_1}{e_2}, & \text{if } \frac{x_1}{y_1} > \frac{x_2}{y_2}, \end{cases} \tag{A.4}
$$

where $e_i = \frac{y_i}{x_i + y_i}$, $i = 1, 2$.

Proof: From (A.2),

- if $\frac{x_1}{y_1} = \frac{x_2}{y_2}$,

$$
\begin{aligned}
\lim_{N\to\infty} Q(x_1, y_1, x_2, y_2, N) &= \lim_{N\to\infty} \frac{x_1(x_1 + x_2)(y_1 + y_2)}{(x_1 + y_1)[(x_1 + x_2)(y_1 + y_2) + x_2 y_1 (x_1 + x_2 + y_1 + y_2) N]} \\
&= 0;
\end{aligned} \tag{A.5}
$$

- if $\frac{x_1}{y_1} < \frac{x_2}{y_2}$, then

$$
\beta = \frac{(x_1 + x_2 + y_1 + y_2)(x_1 y_2 - x_2 y_1)}{(x_1 + x_2)(y_1 + y_2)} < 0. \tag{A.6}
$$

  Therefore,

$$
\lim_{N\to\infty} Q(x_1, y_1, x_2, y_2, N) = \lim_{N\to\infty} \frac{(1-e_1)(1-\phi)}{1 - \phi e^{-\beta N}} = 0, \tag{A.7}
$$

  where $\phi = \frac{e_1(1-e_2)}{e_2(1-e_1)}$;

- if $\frac{x_1}{y_1} < \frac{x_2}{y_2}$, then $\beta > 0$, and, therefore,

$$\begin{aligned}
\lim_{N\to\infty} Q(x_1, y_1, x_2, y_2, N) &= \lim_{N\to\infty} \frac{(1-e_1)(1-\phi)}{1-\phi e^{-\beta N}} \\
&= (1-e_1)(1-\phi) \\
&= (1-e_1)\Big(1 - \frac{e_1(1-e_2)}{e_2(1-e_1)}\Big) \\
&= (1-e_1) - \frac{e_1(1-e_2)}{e_2} \\
&= 1 - e_1 - \frac{e_1}{e_2} + e_1 \\
&= 1 - \frac{e_1}{e_2}.
\end{aligned} \tag{A.8}$$

■

Lemma A.2 *Let $e_j = \min_{1\leqslant i\leqslant M} e_i$. Then, for $N_i = \infty$, $i = 0, 1, \ldots, M-1$, the unique solution of (A.1) is given by*

$$\begin{aligned}
e_i^f &= \begin{cases} e_i, & \text{if } i < j, \\ e_j, & \text{if } i \geqslant j, \end{cases} \\
e_i^b &= \begin{cases} e_j, & \text{if } i \leqslant j, \\ e_i, & \text{if } i > j, \end{cases}
\end{aligned} \tag{A.9}$$

$$\begin{aligned}
\lambda_i^f &= (\lambda_i + \mu_i)(1 - e_i^f),\ \mu_i^f = (\lambda_i + \mu_i)e_i^f, \\
\lambda_i^b &= (\lambda_i + \mu_i)(1 - e_i^b),\ \mu_i^b = (\lambda_i + \mu_i)e_i^b.
\end{aligned} \tag{A.10}$$

Proof: To prove (A.9) and (A.10), we show that it is the solution of (A.1) and then comment its uniqueness.

Since $e_j = \min_{1\leqslant i\leqslant M} e_i$, i.e., $e_j \leqslant e_i$, $\forall i = 0, 1, \ldots, M$, we have

$$\frac{1-e_j}{e_j} \geqslant \frac{1-e_i}{e_i},\ \forall i = 0, 1, \ldots, M. \tag{A.11}$$

- If $i < j$, then based on (A.4), (A.9), (A.10), and (A.11), we have

$$\begin{aligned}
&Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) \\
&= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}^f), (\lambda_{i-1} + \mu_{i-1})e_{i-1}^f, (\lambda_i + \mu_i)(1 - e_i^b), (\lambda_i + \mu_i)e_i^b, \infty] \\
&= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}), (\lambda_{i-1} + \mu_{i-1})e_{i-1}, (\lambda_i + \mu_i)(1 - e_j), (\lambda_i + \mu_i)e_j, \infty] \\
&= 0
\end{aligned} \tag{A.12}$$

and

$$\begin{aligned}
&Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) \\
&= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}^b), (\lambda_{i+1} + \mu_{i+1})e_{i+1}^b, (\lambda_i + \mu_i)(1 - e_i^f), (\lambda_i + \mu_i)e_i^f, \infty] \\
&= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_j), (\lambda_{i+1} + \mu_{i+1})e_j, (\lambda_i + \mu_i)(1 - e_i), (\lambda_i + \mu_i)e_i, \infty] \\
&= 1 - \frac{e_j}{e_i}.
\end{aligned} \tag{A.13}$$

Thus, for the left- and right-hand sides of the first equation of (A.1), we have, respectively,

$$\mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda_i + \mu_i)e_i = \mu_i \tag{A.14}$$

and

$$\mu_i - \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \mu_i, \tag{A.15}$$

implying that (A.9) and (A.10) solve the first equation of (A.1) for $i < j$. Similarly, for the left- and right-hand sides of the second equation of (A.1), we have,

$$\lambda_i^f = (\lambda_i + \mu_i)(1 - e_i^f) = (\lambda_i + \mu_i)(1 - e_i) = \lambda_i \tag{A.16}$$

and

$$\lambda_i + \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \lambda_i, \tag{A.17}$$

implying that (A.9) and (A.10) solve the second equation of (A.1) for $i < j$. For the third equation of (A.1), the left- and right-hand sides are respectively

$$\mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda_i + \mu_i)e_j \tag{A.18}$$

and

$$\mu_i - \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \mu_i \frac{e_j}{e_i} = (\lambda_i + \mu_i)e_j, \tag{A.19}$$

implying that (A.9) and (A.10) solve the third equation of (A.1) for $i < j$. As for the last equation of (A.1), the left- and right-hand sides are

$$\lambda_i^b = (\lambda_i + \mu_i)(1 - e_i^b) = (\lambda_i + \mu_i)(1 - e_j) \tag{A.20}$$

and

$$\lambda_i + \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \lambda_i + \mu_i\Big(1 - \frac{e_j}{e_i}\Big) = (\lambda_i + \mu_i)(1 - e_j), \tag{A.21}$$

implying that (A.9) and (A.10) solve the last equation of (A.1) for $i < j$.

- If $i = j$, the two Q-functions in (A.1) are respectively

$$\begin{aligned} &Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) \\ &= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}^f), (\lambda_{i-1} + \mu_{i-1})e_{i-1}^f, (\lambda_i + \mu_i)(1 - e_i^b), (\lambda_i + \mu_i)e_i^b, \infty] \\ &= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}), (\lambda_{i-1} + \mu_{i-1})e_{i-1}, (\lambda_i + \mu_i)(1 - e_j), (\lambda_i + \mu_i)e_j, \infty] \\ &= 0 \end{aligned} \tag{A.22}$$

and

$$\begin{aligned} &Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) \\ &= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}^b), (\lambda_{i+1} + \mu_{i+1})e_{i+1}^b, (\lambda_i + \mu_i)(1 - e_i^f), (\lambda_i + \mu_i)e_i^f, \infty] \\ &= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}), (\lambda_{i+1} + \mu_{i+1})e_{i+1}, (\lambda_i + \mu_i)(1 - e_j), (\lambda_i + \mu_i)e_j, \infty] \\ &= 0. \end{aligned} \tag{A.23}$$

Thus, the left- and right-hand sides of (A.1) are

$$\begin{aligned}
&\mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda_i + \mu_i)e_j = (\lambda_i + \mu_i)e_i = \mu_i,\\
&\mu_i - \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \mu_i,\\
&\lambda_i^f = (\lambda_i + \mu_i)(1 - e_i^f) = (\lambda_i + \mu_i)(1 - e_j) = (\lambda_i + \mu_i)(1 - e_i) = \lambda_i,\\
&\lambda_i + \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \lambda_i,\\
&\mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda_i + \mu_i)e_j = (\lambda_i + \mu_i)e_i = \mu_i,\\
&\mu_i - \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \mu_i,\\
&\lambda_i^b = (\lambda_i + \mu_i)(1 - e_i^b) = (\lambda_i + \mu_i)(1 - e_j) = (\lambda_i + \mu_i)(1 - e_i) = \lambda_i,\\
&\lambda_i + \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \lambda_i,
\end{aligned} \tag{A.24}$$

implying that (A.1) is solved for $i = j$.

- If $i > j$, the two Q-functions in (A.1) are

$$\begin{aligned}
&Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty)\\
&= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_{i-1}^f), (\lambda_{i-1} + \mu_{i-1})e_{i-1}^f, (\lambda_i + \mu_i)(1 - e_i^b), (\lambda_i + \mu_i)e_i^b, \infty]\\
&= Q[(\lambda_{i-1} + \mu_{i-1})(1 - e_j), (\lambda_{i-1} + \mu_{i-1})e_j, (\lambda_i + \mu_i)(1 - e_i), (\lambda_i + \mu_i)e_i, \infty]\\
&= 1 - \frac{e_j}{e_i}
\end{aligned} \tag{A.25}$$

and

$$\begin{aligned}
&Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty)\\
&= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}^b), (\lambda_{i+1} + \mu_{i+1})e_{i+1}^b, (\lambda_i + \mu_i)(1 - e_i^f), (\lambda_i + \mu_i)e_i^f, \infty]\\
&= Q[(\lambda_{i+1} + \mu_{i+1})(1 - e_{i+1}), (\lambda_{i+1} + \mu_{i+1})e_{i+1}, (\lambda_i + \mu_i)(1 - e_j), (\lambda_i + \mu_i)e_j, \infty]\\
&= 0.
\end{aligned} \tag{A.26}$$

Thus, the left- and right-hand sides of (A.1) are

$$
\begin{aligned}
&\mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda_i + \mu_i)e_j, \\
&\mu_i - \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \mu_i \frac{e_j}{e_i} = (\lambda_i + \mu_i)e_j, \\
&\lambda_i^f = (\lambda_i + \mu_i)(1 - e_i^f) = (\lambda_i + \mu_i)(1 - e_j), \\
&\lambda_i + \mu_i Q(\lambda_{i-1}^f, \mu_{i-1}^f, \lambda_i^b, \mu_i^b, \infty) = \lambda_i + \mu_i\Big(1 - \frac{e_j}{e_i}\Big) = (\lambda_i + \mu_i)(1 - e_j), \\
&\mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda_i + \mu_i)e_i = \mu_i, \\
&\mu_i - \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \mu_i, \\
&\lambda_i^b = (\lambda_i + \mu_i)(1 - e_i^b) = (\lambda_i + \mu_i)(1 - e_i) = \lambda_i, \\
&\lambda_i + \mu_i Q(\lambda_{i+1}^b, \mu_{i+1}^b, \lambda_i^f, \mu_i^f, \infty) = \lambda_i,
\end{aligned}
\tag{A.27}
$$

which also implies that (A.1) is solved.

As far as the uniqueness of (A.9) and (A.10) is concerned, it follows directly from Theorem 11.4 of [3]. ■

Lemma A.3 *In synchronous exponential two-machine lines defined by assumption (a)-(e) in [3], if* $e_1 < e_2$*,*

$$\lim_{N\to\infty} WIP = \frac{e_1}{\tau}\Big(\frac{e_1}{\mu_1} + \frac{e_2}{\mu_2}\Big)\Big(\frac{1 - e_2}{e_2 - e_1}\Big). \tag{A.28}$$

Proof: From the proof of Theorem 11.3 in [3], we know that

$$\lim_{N\to\infty} WIP = \lim_{N\to\infty} \frac{D_5}{D_2 + D_3} = \lim_{N\to\infty} \frac{\frac{D_2}{-K}}{D_2 + D_3}, \tag{A.29}$$

where $\lim_{N\to\infty} D_2 = -\frac{2 + D_1 + \frac{1}{D_1}}{K}$, $D_1 = \frac{\mu_1 + \mu_2}{\lambda_1 + \lambda_2}$, $K = \frac{(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)(\lambda_2\mu_1 - \lambda_1\mu_2)}{(\lambda_1 + \lambda_2)(\mu_1 + \mu_2)}$, $D_3 = \frac{(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)(\lambda_2 + \mu_1) + \lambda_1\mu_2 - \lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)}$. Thus, we have

$$
\begin{aligned}
\lim_{N\to\infty}\frac{1}{WIP} &= \lim_{N\to\infty} -K\Big(\frac{D_3}{D_2}+1\Big) = K^2\frac{D_3}{2+D_1+\frac{1}{D_1}} - K\\
&= \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)^2(\lambda_2\mu_1-\lambda_1\mu_2)^2}{(\lambda_1+\lambda_2)^2(\mu_1+\mu_2)^2}\frac{\frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)+\lambda_1\mu_2-\lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}}{2+\frac{\mu_1+\mu_2}{\lambda_1+\lambda_2}+\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}}\\
&\quad - \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2\mu_1-\lambda_1\mu_2)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}\\
&= \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)^2(\lambda_2\mu_1-\lambda_1\mu_2)^2\frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)+\lambda_1\mu_2-\lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)(\lambda_1+\lambda_2+\mu_1+\mu_2)^2}\\
&\quad - \frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2\mu_1-\lambda_1\mu_2)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}\\
&= \frac{\lambda_2\mu_1-\lambda_1\mu_2}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}\Big[(\lambda_2\mu_1-\lambda_1\mu_2)\frac{(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)+\lambda_1\mu_2-\lambda_2\mu_1}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\\
&\quad -(\lambda_1+\lambda_2+\mu_1+\mu_2)\Big]\\
&= \frac{\lambda_2\mu_1-\lambda_1\mu_2}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big[(\lambda_2\mu_1-\lambda_1\mu_2)(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)\\
&\quad -(\lambda_1\mu_2-\lambda_2\mu_1)^2-\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)^2\Big]\\
&= \frac{\lambda_2\mu_1-\lambda_1\mu_2}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big[(\lambda_2\mu_1-\lambda_1\mu_2)(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)\\
&\quad -(\lambda_1\mu_2-\lambda_2\mu_1)^2-\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)\\
&\quad -\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_1+\mu_2)\Big]\\
&= \frac{\lambda_2\mu_1-\lambda_1\mu_2}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big[-\lambda_1\mu_2(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_2+\mu_1)\\
&\quad -(\lambda_1\mu_2-\lambda_2\mu_1)^2-\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)(\lambda_1+\mu_2)\Big]\\
&= \frac{\lambda_1\mu_2-\lambda_2\mu_1}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}\Big[(\lambda_1+\mu_2)+\frac{\lambda_1\mu_2(\lambda_2+\mu_1)}{\lambda_2\mu_1}+\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big]\\
&= \frac{(\lambda_1+\mu_1)(\lambda_2+\mu_2)(e_2-e_1)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}\Big[(\lambda_1+\mu_2)+\frac{\lambda_1\mu_2(\lambda_2+\mu_1)}{\lambda_2\mu_1}+\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\lambda_2\mu_1(\lambda_1+\lambda_2+\mu_1+\mu_2)}\Big]\\
&= \frac{(\lambda_2+\mu_2)(e_2-e_1)}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)(\lambda_2+\mu_2)(1-e_2)e_1}\Big[\lambda_2\mu_1(\lambda_1+\mu_2)+\lambda_1\mu_2(\lambda_2+\mu_1)\\
&\quad +\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\lambda_1+\lambda_2+\mu_1+\mu_2}\Big]\\
&= \frac{(\lambda_2+\mu_2)(e_2-e_1)}{e_1(1-e_2)}\Big[1+\frac{1}{(\lambda_1+\lambda_2)(\mu_1+\mu_2)(\lambda_2+\mu_2)}\Big(\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\lambda_1+\lambda_2+\mu_1+\mu_2}-\lambda_2^2(\mu_1+\mu_2)\\
&\quad -\mu_2^2(\lambda_1+\lambda_2)\Big)\Big]\\
&= \frac{(\lambda_2+\mu_2)(e_2-e_1)}{e_1(1-e_2)}\Big[1-\frac{\lambda_2+\mu_2}{\lambda_1+\lambda_2+\mu_1+\mu_2}\Big] = \frac{(\lambda_1+\mu_1)(\lambda_2+\mu_2)(e_2-e_1)}{(\lambda_1+\lambda_2+\mu_1+\mu_2)e_1(1-e_2)}\\
&= \frac{\mu_1\mu_2(e_2-e_1)}{(\mu_1e_2+\mu_2e_1)e_1(1-e_2)}.
\end{aligned}
\tag{A.30}
$$

Therefore,

$$\lim_{N\to\infty} WIP = \frac{(\mu_1 e_2 + \mu_2 e_1)e_1(1-e_2)}{\mu_1\mu_2(e_2-e_1)} = e_1\Big(\frac{e_1}{\mu_1} + \frac{e_2}{\mu_2}\Big)\Big(\frac{1-e_2}{e_2-e_1}\Big). \tag{A.31}$$

In [3], (A.29) is derived for $\tau = 1$. For general τ, (A.28) follows (see proofs for Chapter 11 in Chapter 20 of [3]). ■

Proof of Proposition 3.1: For the synchronous exponential production line defined by assumptions (i)-(v) with $\lambda_i = \lambda$, $\mu_i = \mu$, $i = 1, 2, \ldots, M$ and $e_0 < e$, based on Lemma A.2 we obtain

$$\begin{aligned}
&e_0^f = e_0,\ \mu_0^f = (\lambda_0+\mu_0)e_0^f = (\lambda_0+\mu_0)e_0 = \mu_0,\\
&e_i^f = e_0,\ \mu_i^f = (\lambda_i+\mu_i)e_i^f = (\lambda+\mu)e_0,\ i = 1,2,\ldots,M-1,\\
&e_i^b = e,\ \mu_i^b = (\lambda_i+\mu_i)e_i^b = (\lambda+\mu)e = \mu,\ i = 1,2,\ldots,M,
\end{aligned} \tag{A.32}$$

which, using Lemma A.3, implies that the occupancy of each buffer is

$$\begin{aligned}
\widehat{WIP}_0 &= \frac{e_0^f}{\tau}\Big(\frac{e_0^f}{\mu_0^f} + \frac{e_1^b}{\mu_1^b}\Big)\Big(\frac{1-e_1^b}{e_1^b-e_0^f}\Big)\\
&= \frac{e_0}{\tau}\Big(\frac{e_0}{\mu_0} + \frac{e}{\mu}\Big)\Big(\frac{1-e}{e-e_0}\Big)
\end{aligned} \tag{A.33}$$

and

$$\begin{aligned}
\widehat{WIP}_i &= \frac{e_i^f}{\tau}\Big(\frac{e_i^f}{\mu_i^f} + \frac{e_{i+1}^b}{\mu_{i+1}^b}\Big)\Big(\frac{1-e_{i+1}^b}{e_{i+1}^b-e_i^f}\Big)\\
&= \frac{e_0}{\tau}\Big(\frac{1}{\lambda+\mu} + \frac{e}{\mu}\Big)\Big(\frac{1-e}{e-e_0}\Big)\\
&= \frac{2e_0 e}{\mu\tau}\Big(\frac{1-e}{e-e_0}\Big),\ i = 1,2,\ldots,M-1.
\end{aligned} \tag{A.34}$$

Thus, taking into account that

$$\widehat{TP} = \frac{e_0}{\tau}, \tag{A.35}$$

and using Little's law, from (A.33) and (A.34) we obtain the lead time in buffers

$$\widehat{LT}_{buffer} = \frac{\sum_{i=0}^{M-1}\widehat{WIP}_i}{\widehat{TP}} = \Big[\frac{e_0}{\mu_0} + (2M-1)\frac{e}{\mu}\Big]\Big(\frac{1-e}{e-e_0}\Big). \tag{A.36}$$

Considering the lead time on M machines is $M\tau$, we obtain (3.1). ■

Proof of Proposition 3.2: Let

$$f(\rho) := \alpha\widehat{lt}(\rho) = \alpha\Big(1 + \frac{1}{M\tau}\Big(\frac{\rho}{\mu_0} + \frac{2M-1}{\mu}\Big)\Big(\frac{1-e}{1-\rho}\Big)\Big). \tag{A.37}$$

Then

$$\begin{aligned} f'(\rho) &= \frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu}\Big)\frac{1-e}{(1-\rho)^2}, \\ f''(\rho) &= \frac{2\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu}\Big)\frac{1-e}{(1-\rho)^3}, \end{aligned}$$

and, therefore,

$$\kappa(f(\rho)) = \frac{\left|f''_{\rho\rho}\right|}{(1+f'^2_\rho)^{\frac{3}{2}}} = \frac{\frac{2\alpha}{M\tau}(\frac{1}{\mu_0} + \frac{2M-1}{\mu})(1-e)}{\Big[(1-\rho)^2 + \frac{\alpha^2}{M^2\tau^2}(\frac{1}{\mu_0} + \frac{2M-1}{\mu})^2\frac{(1-e)^2}{(1-\rho)^2}\Big]^{\frac{3}{2}}}. \tag{A.38}$$

Since

$$\rho_{knee} = \arg\max_{\rho} \kappa(f(\rho)), \tag{A.39}$$

from (A.38) we obtain (3.12). Clearly, when M tends to infinity, (3.12) becomes (3.13). ■

Proof of Proposition 3.3: Similar to the proof of Proposition 3.1, with the only difference that, instead of Equation (A.32), we have

$$\begin{aligned} &e_i^f = e_0,\ \mu_i^f = (\lambda_i + \mu_i)e_i^f = (\lambda_i + \mu_i)e_0,\ i = 0, 1, \ldots, M-1, \\ &e_i^b = e_i,\ \mu_i^b = (\lambda_i + \mu_i)e_i^b = (\lambda_i + \mu_i)e_i = \mu_i,\ i = 1, 2, \ldots, M \end{aligned} \tag{A.40}$$

and, therefore,

$$\begin{aligned} \widehat{WIP}_0 &= \frac{e_0^f}{\tau}\Big(\frac{e_0^f}{\mu_0^f} + \frac{e_1^b}{\mu_1^b}\Big)\Big(\frac{1-e_1^b}{e_1^b - e_0^f}\Big) \\ &= \frac{e_0}{\tau}\Big(\frac{e_0}{\mu_0} + \frac{e_1}{\mu_1}\Big)\Big(\frac{1-e_1}{e_1 - e_0}\Big), \end{aligned} \tag{A.41}$$

$$
\begin{aligned}
\widehat{WIP}_i &= \frac{e_i^f}{\tau}\Big(\frac{e_i^f}{\mu_i^f} + \frac{e_{i+1}^b}{\mu_{i+1}^b}\Big)\Big(\frac{1-e_{i+1}^b}{e_{i+1}^b - e_i^f}\Big) \\
&= \frac{e_0}{\tau}\Big(\frac{1}{\lambda_i + \mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big) \\
&= \frac{e_0}{\tau}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big),\ i = 1, 2, \ldots, M-1.
\end{aligned}
\tag{A.42}
$$

■

Proof of Proposition 3.4: From (3.16) and (3.4), we obtain

$$
\widehat{lt} = 1 + \frac{1}{M\tau}\sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big). \tag{A.43}
$$

Thus,

$$
\begin{aligned}
\widehat{lt} &\leqslant 1 + \frac{1}{M\tau}\sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{min}}{e_{min}-e_0}\Big) \\
&\leqslant 1 + \frac{1}{M\tau}\Big(\frac{e_0}{\mu_0} + (2M-1)\frac{e_{max}}{\mu_{min}}\Big)\Big(\frac{1-e_{min}}{e_{min}-e_0}\Big) \\
&= 1 + \frac{1}{\tau}\Big(\frac{\rho_{max}}{M\mu_0} + \frac{2M-1}{M\mu_{min}}\frac{e_{max}}{e_{min}}\Big)\Big(\frac{1-e_{min}}{1-\rho_{max}}\Big).
\end{aligned}
\tag{A.44}
$$

■

Proof of Proposition 3.5: Similar to the proof of Proposition 3.2, with the only difference that

$$
\begin{aligned}
f(\rho_{max}) &:= \alpha\widehat{lt}(\rho_{max}) = \alpha\left(1 + \frac{1}{M\tau}\Big(\frac{\rho_{max}}{\mu_0} + \frac{2M-1}{\mu_{min}}\frac{e_{max}}{e_{min}}\Big)\Big(\frac{1-e_{min}}{1-\rho_{max}}\Big)\right), \\
f'(\rho_{max}) &= \frac{\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu_{min}}\frac{e_{max}}{e_{min}}\Big)\frac{1-e_{min}}{(1-\rho_{max})^2}, \\
f''(\rho_{max}) &= \frac{2\alpha}{M\tau}\Big(\frac{1}{\mu_0} + \frac{2M-1}{\mu_{min}}\frac{e_{max}}{e_{min}}\Big)\frac{1-e_{min}}{(1-\rho_{max})^3},
\end{aligned}
$$

and, therefore, we obtain (3.23). Clearly, when M tends to infinity, (3.23) becomes (3.24). ■

Proof of Proposition 3.6: Let

$$
f(\rho_{max}) := \alpha\left\{1 + \frac{1}{M\tau}\sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big)\right\}. \tag{A.45}
$$

To prove the proposition, we need to prove that

$$\kappa(f(\rho_{max})) = \frac{|f''(\rho_{max})|}{(1 + f'^2(\rho_{max}))^{\frac{3}{2}}} \tag{A.46}$$

is an increasing function of $\rho_{max} \in (0, \bar{\bar{\rho}}_{\infty,knee}]$, where $\bar{\bar{\rho}}_{\infty,knee}$ is defined in (3.24). In other words, we need to prove

$$\kappa'(f(\rho_{max})) = \frac{f'''(\rho_{max})(1 + f'^2(\rho_{max})) - 3f'(\rho_{max})f''^2(\rho_{max})}{(1 + f'^2(\rho_{max}))^{\frac{5}{2}}} \geqslant 0 \tag{A.47}$$

for all $\rho_{max} \in (0, \bar{\bar{\rho}}_{\infty,knee}]$, where

$$\begin{aligned}
f'(\rho_{max}) &= \frac{\alpha}{M\tau e_{min}}\Big[\Big(\frac{e_1}{\mu_0} + \frac{e_1}{\mu_1}\Big)\frac{1-e_1}{(\frac{e_1}{e_{min}} - \rho_{max})^2} + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\frac{1-e_{i+1}}{(\frac{e_{i+1}}{e_{min}} - \rho_{max})^2}\Big],\\
f''(\rho_{max}) &= \frac{2\alpha}{M\tau e_{min}}\Big[\Big(\frac{e_1}{\mu_0} + \frac{e_1}{\mu_1}\Big)\frac{1-e_1}{(\frac{e_1}{e_{min}} - \rho_{max})^3} + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\frac{1-e_{i+1}}{(\frac{e_{i+1}}{e_{min}} - \rho_{max})^3}\Big],\\
f'''(\rho_{max}) &= \frac{6\alpha}{M\tau e_{min}}\Big[\Big(\frac{e_1}{\mu_0} + \frac{e_1}{\mu_1}\Big)\frac{1-e_1}{(\frac{e_1}{e_{min}} - \rho_{max})^4} + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\frac{1-e_{i+1}}{(\frac{e_{i+1}}{e_{min}} - \rho_{max})^4}\Big].
\end{aligned} \tag{A.48}$$

Let

$$\begin{aligned}
&\gamma_1 = \Big(\frac{e_1}{\mu_0} + \frac{e_1}{\mu_1}\Big)(1 - e_1),\ \gamma_{i+1} = \Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)(1 - e_{i+1}),\ i = 1, 2, \ldots, M-1,\\
&\eta_i = \frac{1}{\frac{e_i}{e_{min}} - \rho_{max}},\ i = 1, 2, \ldots, M.
\end{aligned} \tag{A.49}$$

Then, proving (A.47) implies proving

$$\frac{1}{2}\Big(\sum_{i=1}^{M}\gamma_i\eta_i^4\Big)\Big[\Big(\frac{M\tau e_{min}}{\alpha}\Big)^2 + \Big(\sum_{i=1}^{M}\gamma_i\eta_i^2\Big)^2\Big] \geqslant \Big(\sum_{i=1}^{M}\gamma_i\eta_i^2\Big)\Big(\sum_{i=1}^{M}\gamma_i\eta_i^3\Big)^2. \tag{A.50}$$

In the following, we will first prove

$$\frac{M\tau e_{min}}{\alpha} \geqslant \sum_{i=1}^{M}\gamma_i\eta_i^2 \tag{A.51}$$

for all $\rho_{max} \in (0, \bar{\bar{\rho}}_{\infty,knee}]$ and then

$$\Big(\sum_{i=1}^{M} \gamma_i \eta_i^4\Big)\Big(\sum_{i=1}^{M} \gamma_i \eta_i^2\Big) \geqslant \Big(\sum_{i=1}^{M} \gamma_i \eta_i^3\Big)^2. \tag{A.52}$$

Based on (3.24) and considering that $\mu_0 \geqslant \mu_{min}$, we obtain

$$\frac{M\tau e_{min}}{\alpha} = \frac{2Me_{max}(1-e_{min})}{\mu_{min}(1-\bar{\bar{\rho}}_{\infty,knee})^2} \geqslant \sum_{i=1}^{M} \frac{\gamma_i}{(1-\bar{\bar{\rho}}_{\infty,knee})^2} \geqslant \sum_{i=1}^{M} \gamma_i \eta_i^2,\ \forall \rho_{max} \in (0, \bar{\bar{\rho}}_{\infty,knee}]. \tag{A.53}$$

As for (A.52), let vectors $\boldsymbol{v}_1 = (\sqrt{\gamma_1}\eta_1^2, \sqrt{\gamma_2}\eta_2^2, \ldots, \sqrt{\gamma_M}\eta_M^2)$ and $\boldsymbol{v}_2 = (\sqrt{\gamma_1}\eta_1, \sqrt{\gamma_2}\eta_2, \ldots, \sqrt{\gamma_M}\eta_M)$. Since for vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, we have

$$\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = |\boldsymbol{v}_1| \cdot |\boldsymbol{v}_2| \cos\theta \leqslant |\boldsymbol{v}_1| \cdot |\boldsymbol{v}_2|, \tag{A.54}$$

where $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2$ is the inner product of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ and θ is the angle between these two vectors. Thus, we obtain

$$\sum_{i=1}^{M} \gamma_i \eta_i^3 \leqslant \sqrt{\Big(\sum_{i=1}^{M} \gamma_i \eta_i^4\Big)\Big(\sum_{i=1}^{M} \gamma_i \eta_i^2\Big)}, \tag{A.55}$$

which completes the proof. ■

Proof of Proposition 4.1: From (3.5), we can see that $\widehat{lt}$ is an increasing function of ρ. Since $0 < \rho < 1$, this implies that

$$\widehat{lt} > 1 + (1-e)\frac{2M-1}{M\mu\tau}, \tag{A.56}$$

i.e.,

$$\widehat{lt} > 1 + (1-e)\frac{2M-1}{M}\frac{T_{down}}{\tau}. \tag{A.57}$$

Clearly, for sufficiently large M, the above inequality becomes

$$\widehat{lt}_{\infty} > 1 + 2(1-e)\frac{T_{down}}{\tau}. \tag{A.58}$$

■

Proof of Proposition 4.2: From (3.5) it follows that

$$\hat{\rho}^* = 1 - \frac{\mu + (2M-1)\mu_0}{M\mu\mu_0\tau(lt_d - 1) + \mu(1-e)}(1-e), \tag{A.59}$$

which implies that (4.2) holds. As for (4.3), it follows immediately from the proof of Proposition 3.1. ■

Proof of Proposition 4.3: Re-write (3.16) as

$$\begin{aligned}\widehat{LT} - M\tau &= \frac{e_0}{\mu_0}\Big(\frac{1-e_1}{e_1-e_0}\Big) + \frac{e_1}{\mu_1}\Big(\frac{1-e_1}{e_1-e_0}\Big) + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big) \\ &= \frac{1-e_1}{\mu_0}\Big(\frac{e_1}{e_1-e_0} - 1\Big) + \frac{e_1}{\mu_1}\Big(\frac{1-e_1}{e_1-e_0}\Big) + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big),\end{aligned} \tag{A.60}$$

and taking into account that $0 < e_0 < \min\limits_{1\leqslant i\leqslant M} e_i$, we observe that the right-hand side of (A.60) is a monotonically increasing function of e_0. Thus,

$$\widehat{LT} - M\tau > \sum_{i=1}^{M}\frac{1-e_i}{\mu_i} + \sum_{i=1}^{M-1}\frac{e_i(1-e_{i+1})}{\mu_i e_{i+1}}, \tag{A.61}$$

i.e., the first inequality of (4.9) holds.

As for the second and the third inequalities, from (3.19) and (3.20), respectively, we can see that $\widehat{\overline{\overline{lt}}}$ and $\widehat{\overline{\overline{lt}}}_\infty$ are increasing functions of $0 < \rho_{max} < 1$, which implies that the last two inequalities of (4.9) hold.

■

Proof of Proposition 4.4: Under the assumptions of Proposition 3.3, for any desired lead time LT_d satisfying (4.9), the release rate $\hat{e}_0^*$ that ensures this lead time is a real root less than $\min\limits_{1\leqslant i\leqslant M} e_i$ of the equation

$$LT_d = M + \frac{1}{\tau}\sum_{i=0}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big) \tag{A.62}$$

or

$$(LT_d - M)\tau = \frac{1-e_1}{\mu_0}\Big(\frac{e_1}{e_1-e_0} - 1\Big) + \frac{e_1}{\mu_1}\Big(\frac{1-e_1}{e_1-e_0}\Big) + \sum_{i=1}^{M-1}\Big(\frac{e_i}{\mu_i} + \frac{e_{i+1}}{\mu_{i+1}}\Big)\Big(\frac{1-e_{i+1}}{e_{i+1}-e_0}\Big). \tag{A.63}$$

Since the right-hand side of (A.63) is monotonically increasing with e_0 when $0 < e_0 < \min\limits_{1\leqslant i\leqslant M} e_i$, equation (A.63) has a unique real solution less than $\min\limits_{1\leqslant i\leqslant M} e_i$ ensuring the desired lead time LT_d. Multiplying (A.63) by $\prod\limits_{j=0}^{M-1}(e_{j+1} - e_0)$, we have

$$(LT_d - M)\tau \prod_{i=0}^{M-1}(e_{i+1}-e_0) = (1-e_1)\Big(\frac{e_0}{\mu_0}+\frac{e_1}{\mu_1}\Big)\prod_{i=1}^{M-1}(e_{i+1}-e_0) + \sum_{i=1}^{M-1}\Big((1-e_{i+1})\Big(\frac{e_i}{\mu_i}+\frac{e_{i+1}}{\mu_{i+1}}\Big)\prod_{j=0,j\neq i}^{M-1}(e_{j+1}-e_0)\Big), \tag{A.64}$$

i.e.,

$$(LT_d - M)\tau \prod_{i=0}^{M-1}(e_{i+1}-e_0) - (1-e_1)\Big(\frac{e_0}{\mu_0}+\frac{e_1}{\mu_1}\Big)\prod_{i=1}^{M-1}(e_{i+1}-e_0) - \sum_{i=1}^{M-1}\Big((1-e_{i+1})\Big(\frac{e_i}{\mu_i}+\frac{e_{i+1}}{\mu_{i+1}}\Big)\prod_{j=0,j\neq i}^{M-1}(e_{j+1}-e_0)\Big) = 0. \tag{A.65}$$

In other words, for any desired lead time LT_d satisfying (4.9), the release rate $\hat{e}_0^*$ that ensures this lead time is the unique real root less than $\min\limits_{1\leqslant i\leqslant M} e_i$ of the M-th order polynomial equation (4.10).

The statements on $\widehat{PR}^*$ and $\widehat{WIP}^*$ follow from the proof of Proposition 3.3. ■

Proof of Proposition 4.5: We first prove (4.14) and (4.15) and then comment on (4.16).

From (3.19) and (3.20), it follows, respectively, that

$$\bar{\hat{\rho}}^*_{max} = 1 - \frac{\mu_{min} + (2M-1)\mu_0\frac{e_{max}}{e_{min}}}{M\mu_{min}\mu_0\tau(lt_d - 1) + \mu_{min}(1-e_{min})}(1-e_{min}) \tag{A.66}$$

and

$$\bar{\hat{\rho}}^*_{max,\infty} = 1 - \frac{2(1-e_{min})\frac{e_{max}}{e_{min}}}{\tau\mu_{min}(lt_d - 1)}. \tag{A.67}$$

Thus, we obtain (4.14) and (4.15).

Clearly, if $\mu_0 \geqslant \mu_{min}$, then (3.21) holds, which implies that (4.16) holds. ■

References

[1] S. Biller, S. M. Meerkov, and C.-B. Yan, "Raw material release rates to ensure desired production lead time in Bernoulli serial lines," *International Journal of Production Research*, vol. 51, no. 14, pp. 4349–4364, 2013.

[2] S. M. Meerkov and C.-B. Yan, "Cellular production lines with asymptotically reliable Bernoulli machines: Lead time analysis and control," 2013, submitted to *IEEE Transactions on Automation Science and Engineering*.

[3] J. Li and S. M. Meerkov, *Production Systems Engineering*. New York: Springer, 2009.

[4] J. H. Blackstone Jr., D. T. Phillips, and G. L. Hogg, "A state-of-the-art survey of dispatching rules for manufacturing job shop operations," *The International Journal of Production Research*, vol. 20, no. 1, pp. 27–45, 1982.

[5] R. T. Barrett and S. N. Kadipasaoglu, "Dispatching rules for a dynamic flow shop," *Production and Inventory Management Journal*, vol. 31, no. 1, pp. 54–58, 1990.

[6] A. M. Waikar, B. R. Sarker, and A. M. Lal, "A comparative study of some priority dispatching rules under different shop loads," *Production Planning & Control*, vol. 6, no. 4, pp. 301–310, 1995.

[7] Y.-F. Hung and I.-R. Chen, "A simulation study of dispatch rules for reducing flow times in semiconductor wafer fabrication," *Production Planning & Control*, vol. 9, no. 7, pp. 714–722, 1998.

[8] C. Rajendran and O. Holthaus, "A comparative study of dispatching rules in dynamic flowshops and jobshops," *European Journal of Operational Research*, vol. 116, no. 1, pp. 156–170, 1999.

[9] M. L. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 4th ed. Springer, 2012.

[10] A. Hübl, H. Jodlbauer, and K. Altendorfer, “Influence of dispatching rules on average production lead time for multi-stage production systems,” *International Journal of Production Economics*, vol. 144, no. 2, pp. 479–484, 2013.

[11] Y. Sugimori, K. Kusunoki, F. Cho, and S. Uchikawa, “Toyota production system and kanban system materialization of just-in-time and respect-for-human system,” *International Journal of Production Research*, vol. 15, no. 6, pp. 553–564, 1977.

[12] G. R. Bitran and L. Chang, “A mathematical programming approach to a deterministic kanban system,” *Management Science*, vol. 33, no. 4, pp. 427–441, 1987.

[13] L. J. Krajewski, B. E. King, L. P. Ritzman, and D. S. Wong, “Kanban, MRP, and shaping the manufacturing environment,” *Management Science*, vol. 33, no. 1, pp. 39–57, 1987.

[14] D. Mitra and I. Mitrani, “Analysis of a kanban discipline for cell coordination in production lines. I,” *Management Science*, vol. 36, no. 12, pp. 1548–1566, 1990.

[15] B. J. Berkley, “A review of the kanban production control research literature,” *Production and Operations Management*, vol. 1, no. 4, pp. 393–411, 1992.

[16] S. R. Tayur, “Structural properties and a heuristic for kanban-controlled serial lines,” *Management Science*, vol. 39, no. 11, pp. 1347–1368, 1993.

[17] J. A. Muckstadt and R. T. Sridhar, “A comparison of alternative kanban control mechanisms. I: Background and structural results,” *IIE Transactions*, vol. 27, no. 2, pp. 140–150, 1995.

[18] M. Di Mascolo, Y. Frein, and Y. Dallery, “An analytical method for performance evaluation of kanban controlled production systems,” *Operations Research*, vol. 44, no. 1, pp. 50–64, 1996.

[19] M. S. Akturk and F. Erhun, “An overview of design and operational issues of kanban systems,” *International Journal of Production Research*, vol. 37, no. 17, pp. 3859–3881, 1999.

[20] M. Lage Jr. and M. G. Filho, “Variations of the kanban system: Literature review and classification,” *International Journal of Production Economics*, vol. 125, no. 1, pp. 13–21, 2010.

[21] E. Topan and Z. M. Avsar, "An approximation for kanban controlled assembly systems," *Annals of Operations Research*, vol. 182, no. 1, pp. 133–162, 2011.

[22] J. Matzka, M. Di Mascolo, and K. Furmans, "Buffer sizing of a heijunka kanban system," *Journal of Intelligent Manufacturing*, vol. 23, no. 1, pp. 49–60, 2012.

[23] R. Sato and Y. Khojasteh-Ghamari, "An integrated framework for card-based production control systems," *Journal of Intelligent Manufacturing*, vol. 23, no. 3, pp. 717–731, 2012.

[24] M. L. Spearman, D. L. Woodruff, and W. J. Hopp, "CONWIP: A pull alternative to kanban," *International Journal of Production Research*, vol. 28, no. 5, pp. 879–894, 1990.

[25] I. Duenyas and W. J. Hopp, "Estimating the throughput of an exponential CONWIP assembly system," *Queueing Systems*, vol. 14, no. 1, pp. 135–157, 1993.

[26] S. Gstettner and H. Kuhn, "Analysis of production control systems kanban and CONWIP," *International Journal of Production Research*, vol. 34, no. 11, pp. 3253–3273, 1996.

[27] W. J. Hopp and M. L. Roof, "Setting WIP levels with statistical throughput control (STC) in CONWIP production lines," *International Journal of Production Research*, vol. 36, no. 4, pp. 867–882, 1998.

[28] A. M. Bonvik, Y. Dallery, and S. B. Gershwin, "Approximate analysis of production systems operated by a CONWIP/finite buffer hybrid control policy," *International Journal of Production Research*, vol. 38, no. 13, pp. 2845–2869, 2000.

[29] K. K. Yang, "Managing a flow line with single-kanban, dual-kanban or CONWIP," *Production and Operations Management*, vol. 9, no. 4, pp. 349–366, 2000.

[30] W. G. Gilland, "A simulation study comparing performance of CONWIP and bottleneck-based release rules," *Production Planning & Control*, vol. 13, no. 2, pp. 211–219, 2002.

[31] J. M. Framinan, P. L. Gonzalez, and R. Ruiz-Usano, "The CONWIP production control system: Review and research issues," *Production Planning & Control*, vol. 14, no. 3, pp. 255–265, 2003.

[32] J. Geraghty and C. Heavey, "A comparison of hybrid push/pull and CONWIP/pull production inventory control policies," *International Journal of Production Economics*, vol. 91, no. 1, pp. 75–90, 2004.

[33] H. Jodlbauer and A. Huber, "Service-level performance of MRP, kanban, CONWIP and DBR due to parameter stability and environmental robustness," *International Journal of Production Research*, vol. 46, no. 8, pp. 2179–2195, 2008.

[34] Y. Khojasteh-Ghamari, "A performance comparison between kanban and CONWIP controlled assembly systems," *Journal of Intelligent Manufacturing*, vol. 20, no. 6, pp. 751–760, 2009.

[35] ——, "Developing a framework for performance analysis of a production process controlled by kanban and CONWIP," *Journal of Intelligent Manufacturing*, vol. 23, no. 1, pp. 61–71, 2012.

[36] R. Courant and F. John, *Introduction to Calculus and Analysis*. Springer, 1999, vol. 1.

[37] J. Li and S. M. Meerkov, "On the coefficients of variation of uptime and downtime in manufacturing equipment," *Mathematical Problems in Engineering*, vol. 2005, no. 1, pp. 1–6, 2005.